



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁴ : C12N 15/00, C12P 21/00, 19/34 C12N 9/60	A2	(11) International Publication Number: WO 88/ 02406 (43) International Publication Date: 7 April 1988 (07.04.88)
---	----	---

(21) International Application Number: PCT/US87/02522

(22) International Filing Date: 1 October 1987 (01.10.87)

(31) Priority Application Number: 915,151

(32) Priority Date: 2 October 1986 (02.10.86)

(33) Priority Country: US

(71) Applicant: MASSACHUSETTS INSTITUTE OF TECHNOLOGY [US/US]; 77 Massachusetts Avenue, Cambridge, MA 02139 (US).

(72) Inventors: BACHMAIR, Andreas ; 14 Bidelow Street, Cambridge, MA 02139 (US). FINLEY, Daniel ; 27 Lee Street, Cambridge, MA 02139 (US). VARSHAVSKY, Alexander ; 311 Commonwealth Avenue #30, Boston, MA 02115 (US).

(74) Agents: DeCONTI, Giulio, A., Jr. et al.; Hamilton, Brook, Smith & Reynolds, Two Militia Drive, Lexington, MA 02173 (US).

(81) Designated States: AT (European patent), BE (European patent), CH (European patent), DE (European patent), FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent).

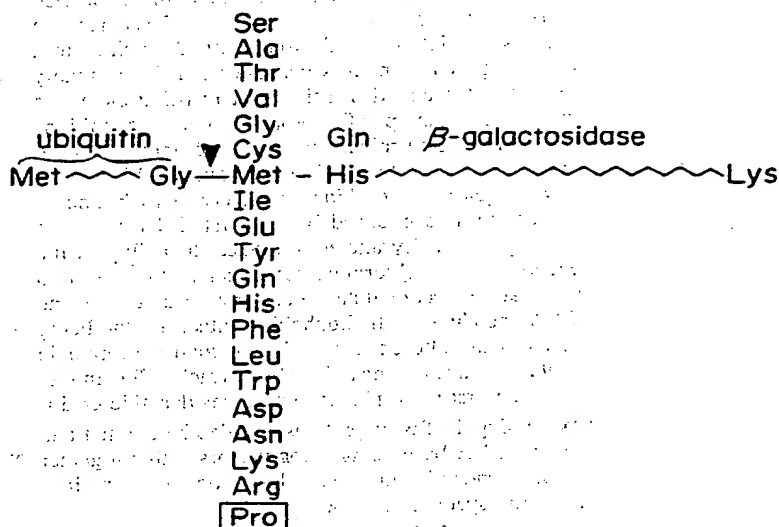
Published

Without international search report and to be republished upon receipt of that report.

(54) Title: METHODS OF REGULATING METABOLIC STABILITY OF PROTEINS

(57) Abstract

Methods of designing or modifying protein structure at the protein or genetic level to produce specified amino termini *in vivo* or *in vitro*. The methods can be used to alter the metabolic stability and other properties of the protein or, alternatively, to artificially generate authentic amino-termini in proteins produced through artificial means. The methods are based upon the introduction of the use of artificial ubiquitin-protein fusions, and the discovery that the *in vivo* half-life of a protein is a function of the amino-terminal amino acid of the protein.



MQIFV...RLRGG—MHGSGAWLLPVSLVKRKTTLAPN...VWCQK

ubiquitin

 β -galactosidase

(76 aa)

(1045 aa)

1. *Pharmaceuticals* (1997) 10, 11. *Pharmaceuticals* (1998) 11, 12. *Pharmaceuticals* (1999) 12, 13. *Pharmaceuticals* (2000) 13, 14. *Pharmaceuticals* (2001) 14, 15. *Pharmaceuticals* (2002) 15, 16. *Pharmaceuticals* (2003) 16, 17. *Pharmaceuticals* (2004) 17, 18. *Pharmaceuticals* (2005) 18, 19. *Pharmaceuticals* (2006) 19, 20. *Pharmaceuticals* (2007) 20, 21. *Pharmaceuticals* (2008) 21, 22. *Pharmaceuticals* (2009) 22, 23. *Pharmaceuticals* (2010) 23, 24. *Pharmaceuticals* (2011) 24, 25. *Pharmaceuticals* (2012) 25, 26. *Pharmaceuticals* (2013) 26, 27. *Pharmaceuticals* (2014) 27, 28. *Pharmaceuticals* (2015) 28, 29. *Pharmaceuticals* (2016) 29, 30. *Pharmaceuticals* (2017) 30, 31. *Pharmaceuticals* (2018) 31, 32. *Pharmaceuticals* (2019) 32, 33. *Pharmaceuticals* (2020) 33, 34. *Pharmaceuticals* (2021) 34, 35. *Pharmaceuticals* (2022) 35, 36. *Pharmaceuticals* (2023) 36, 37. *Pharmaceuticals* (2024) 37, 38. *Pharmaceuticals* (2025) 38, 39. *Pharmaceuticals* (2026) 39, 40. *Pharmaceuticals* (2027) 40, 41. *Pharmaceuticals* (2028) 41, 42. *Pharmaceuticals* (2029) 42, 43. *Pharmaceuticals* (2030) 43, 44. *Pharmaceuticals* (2031) 44, 45. *Pharmaceuticals* (2032) 45, 46. *Pharmaceuticals* (2033) 46, 47. *Pharmaceuticals* (2034) 47, 48. *Pharmaceuticals* (2035) 48, 49. *Pharmaceuticals* (2036) 49, 50. *Pharmaceuticals* (2037) 50, 51. *Pharmaceuticals* (2038) 51, 52. *Pharmaceuticals* (2039) 52, 53. *Pharmaceuticals* (2040) 53, 54. *Pharmaceuticals* (2041) 54, 55. *Pharmaceuticals* (2042) 55, 56. *Pharmaceuticals* (2043) 56, 57. *Pharmaceuticals* (2044) 57, 58. *Pharmaceuticals* (2045) 58, 59. *Pharmaceuticals* (2046) 59, 60. *Pharmaceuticals* (2047) 60, 61. *Pharmaceuticals* (2048) 61, 62. *Pharmaceuticals* (2049) 62, 63. *Pharmaceuticals* (2050) 63, 64. *Pharmaceuticals* (2051) 64, 65. *Pharmaceuticals* (2052) 65, 66. *Pharmaceuticals* (2053) 66, 67. *Pharmaceuticals* (2054) 67, 68. *Pharmaceuticals* (2055) 68, 69. *Pharmaceuticals* (2056) 69, 70. *Pharmaceuticals* (2057) 70, 71. *Pharmaceuticals* (2058) 71, 72. *Pharmaceuticals* (2059) 72, 73. *Pharmaceuticals* (2060) 73, 74. *Pharmaceuticals* (2061) 74, 75. *Pharmaceuticals* (2062) 75, 76. *Pharmaceuticals* (2063) 76, 77. *Pharmaceuticals* (2064) 77, 78. *Pharmaceuticals* (2065) 78, 79. *Pharmaceuticals* (2066) 79, 80. *Pharmaceuticals* (2067) 80, 81. *Pharmaceuticals* (2068) 81, 82. *Pharmaceuticals* (2069) 82, 83. *Pharmaceuticals* (2070) 83, 84. *Pharmaceuticals* (2071) 84, 85. *Pharmaceuticals* (2072) 85, 86. *Pharmaceuticals* (2073) 86, 87. *Pharmaceuticals* (2074) 87, 88. *Pharmaceuticals* (2075) 88, 89. *Pharmaceuticals* (2076) 89, 90. *Pharmaceuticals* (2077) 90, 91. *Pharmaceuticals* (2078) 91, 92. *Pharmaceuticals* (2079) 92, 93. *Pharmaceuticals* (2080) 93, 94. *Pharmaceuticals* (2081) 94, 95. *Pharmaceuticals* (2082) 95, 96. *Pharmaceuticals* (2083) 96, 97. *Pharmaceuticals* (2084) 97, 98. *Pharmaceuticals* (2085) 98, 99. *Pharmaceuticals* (2086) 99, 100. *Pharmaceuticals* (2087) 100, 101. *Pharmaceuticals* (2088) 101, 102. *Pharmaceuticals* (2089) 102, 103. *Pharmaceuticals* (2090) 103, 104. *Pharmaceuticals* (2091) 104, 105. *Pharmaceuticals* (2092) 105, 106. *Pharmaceuticals* (2093) 106, 107. *Pharmaceuticals* (2094) 107, 108. *Pharmaceuticals* (2095) 108, 109. *Pharmaceuticals* (2096) 109, 110. *Pharmaceuticals* (2097) 110, 111. *Pharmaceuticals* (2098) 111, 112. *Pharmaceuticals* (2099) 112, 113. *Pharmaceuticals* (2100) 113, 114. *Pharmaceuticals* (2101) 114, 115. *Pharmaceuticals* (2102) 115, 116. *Pharmaceuticals* (2103) 116, 117. *Pharmaceuticals* (2104) 117, 118. *Pharmaceuticals* (2105) 118, 119. *Pharmaceuticals* (2106) 119, 120. *Pharmaceuticals* (2107) 120, 121. *Pharmaceuticals* (2108) 121, 122. *Pharmaceuticals* (2109) 122, 123. *Pharmaceuticals* (2110) 123, 124. *Pharmaceuticals* (2111) 124, 125. *Pharmaceuticals* (2112) 125, 126. *Pharmaceuticals* (2113) 126, 127. *Pharmaceuticals* (2114) 127, 128. *Pharmaceuticals* (2115) 128, 129. *Pharmaceuticals* (2116) 129, 130. *Pharmaceuticals* (2117) 130, 131. *Pharmaceuticals* (2118) 131, 132. *Pharmaceuticals* (2119) 132, 133. *Pharmaceuticals* (2120) 133, 134. *Pharmaceuticals* (2121) 134, 135. *Pharmaceuticals* (2122) 135, 136. *Pharmaceuticals* (2123) 136, 137. *Pharmaceuticals* (2124) 137, 138. *Pharmaceuticals* (2125) 138, 139. *Pharmaceuticals* (2126) 139, 140. *Pharmaceuticals* (2127) 140, 141. *Pharmaceuticals* (2128) 141, 142. *Pharmaceuticals* (2129) 142, 143. *Pharmaceuticals* (2130) 143, 144. *Pharmaceuticals* (2131) 144, 145. *Pharmaceuticals* (2132) 145, 146. *Pharmaceuticals* (2133) 146, 147. *Pharmaceuticals* (2134) 147, 148. *Pharmaceuticals* (2135) 148, 149. *Pharmaceuticals* (2136) 149, 150. *Pharmaceuticals* (2137) 150, 151. *Pharmaceuticals* (2138) 151, 152. *Pharmaceuticals* (2139) 152, 153. *Pharmaceuticals* (2140) 153, 154. *Pharmaceuticals* (2141) 154, 155. *Pharmaceuticals* (2142) 155, 156. *Pharmaceuticals* (2143) 156, 157. *Pharmaceuticals* (2144) 157, 158. *Pharmaceuticals* (2145) 158, 159. *Pharmaceuticals* (2146) 159, 160. *Pharmaceuticals* (2147) 160, 161. *Pharmaceuticals* (2148) 161, 162. *Pharmaceuticals* (2149) 162, 163. *Pharmaceuticals* (2150) 163, 164. *Pharmaceuticals* (2151) 164, 165. *Pharmaceuticals* (2152) 165, 166. *Pharmaceuticals* (2153) 166, 167. *Pharmaceuticals* (2154) 167, 168. *Pharmaceuticals* (2155) 168, 169. *Pharmaceuticals* (2156) 169, 170. *Pharmaceuticals* (2157) 170, 171. *Pharmaceuticals* (2158) 171, 172. *Pharmaceuticals* (2159) 172, 173. *Pharmaceuticals* (2160) 173, 174. *Ph*

[illegible]

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	ML	Mali
AU	Australia	GA	Gabon	MR	Mauritania
BB	Barbados	GB	United Kingdom	MW	Malawi
BE	Belgium	HU	Hungary	NL	Netherlands
BG	Bulgaria	IT	Italy	NO	Norway
BJ	Benin	JP	Japan	RO	Romania
BR	Brazil	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	LI	Liechtenstein	SN	Senegal
CH	Switzerland	LK	Sri Lanka	SU	Soviet Union
CM	Cameroon	LU	Luxembourg	TD	Chad
DE	Germany, Federal Republic of	MC	Monaco	TG	Togo
DK	Denmark	MG	Madagascar	US	United States of America
FI	Finland				

-1-

METHODS OF REGULATING METABOLIC
STABILITY OF PROTEINS

Background of the Invention

In both bacterial and eukaryotic cells, relatively long-lived proteins, whose half-lives are close to or exceed the cell generation time, coexist with proteins whose half-lives can be less than one percent of the cell generation time. Rates of intracellular protein degradation are a function of the cell's physiological state, and appear to be controlled differentially for individual proteins. In particular, damaged and otherwise abnormal proteins are metabolically unstable *in vivo*. Although the specific functions of selective protein degradation are in most cases still unknown; it is clear that many regulatory proteins are extremely short-lived in vivo. Metabolic instability of such

-2-

proteins allows for rapid adjustment of their intracellular concentrations through regulated changes in rates of their synthesis or degradation. The few instances in which the metabolic instability of an intracellular protein has been shown to be essential for its function include the cII protein of bacteriophage lambda and the HO endonuclease of the yeast Saccharomyces cerevisiae.

Most of the selective turnover of intracellular proteins under normal metabolic conditions is ATP-dependent and (in eukaryotes) nonlysosomal. Recent biochemical and genetic evidence indicates that, in eukaryotes, covalent conjugation of ubiquitin to short-lived intracellular proteins is essential for their selective degradation. The rules which determine whether a given protein is metabolically stable or unstable in vivo were previously unknown.

Summary of the Invention

This invention pertains to methods of engineering the amino terminus of proteins thereby controlling the metabolic stability and other properties of a protein. Further, this invention provides a method for either in vivo or in vitro production of proteins with any of the twenty amino acid residues (or analogs thereof) at the protein's amino-terminus. The invention is based in part upon the striking discovery that the in vivo half-life of

-3-

an intracellular protein is a function of its amino-terminal amino acid residue and upon a novel (and more generally applicable) technique that allows one to generate proteins with specified amino-termini in vivo or in vitro. The invention also pertains to a newly identified protease, ubiquitin-specific processing protease, which has properties that allow one to expose, either in vitro or in vivo, any desired amino acid residue at the amino-terminus of a protein of interest.

The nature of the amino acid exposed at the amino-terminus of an intracellular protein was shown to be one crucial determinant that specifies whether a protein will be long- or short-lived in vivo. Individual amino acids can be categorized as either stabilizing or destabilizing amino acids with respect to the half-life that they confer upon a protein when exposed at the protein's amino-terminus. Destabilizing amino acid residues confer short half-lives, down to a few minutes for some of the destabilizing amino acids. Stabilizing amino acid residues confer long half-lives of many hours. This striking and newly discovered dependency of a protein's half-life on its amino-terminal residue is referred to herein as the N-end rule.

Based upon the N-end rule, the amino-terminus of a protein can thus be designed or altered to change the intracellular half-life of the protein.

and in this way the lifetime and/or activity of the protein in vivo can be regulated. This capability can be exploited for rational protein design in many different contexts. Natural or wild type proteins can be modified to render them more or less resistant to degradation in vivo. The design or alteration of the protein can be done at the protein level or at the genetic (DNA) level. For example, proteins can be modified by chemically altering or engineering the amino-terminus to provide for exposure at the amino-terminus of an amino acid residue of the stabilizing or destabilizing class. At the genetic level, genes encoding proteins can be made to encode an amino acid of the desired class at the amino-terminus so that the expressed protein exhibits a predetermined amino-terminal structure which renders it either metabolically stable or unstable with respect to the N-end rule pathway of proteolytic degradation. Furthermore, proteins can be expressed fused to a "masking" protein sequence which masks the engineered amino-terminus so that when unmasked the protein will exhibit the desired stability or other properties that depend on the nature of the protein's amino-terminal residue. In such constructs, for example, the junction between the two protein sequences can be designed to be cleaved specifically, for instance, by an endoprotease. Endoproteolytic cleavage of the fused sequence unmasks the

-5-

specifically engineered amino-terminus of the protein of interest and subjects the protein to degradation governed by the N-end rule. One specific and new way to engineer the protein's amino-terminus is provided in this invention by the identification of ubiquitin-specific processing protease and determination of its substrate specificity. Using this protease, fusions of ubiquitin with other proteins can be specifically processed either in vitro or in vivo to generate proteins with desired amino-terminal residues.

A different, and also new way to specifically engineer short-lived proteins is provided in this invention by the discovery that ubiquitin-protein fusions, such as ubiquitin-Pro- β -galactosidase, that cannot be efficiently deubiquitinated, are metabolically unstable. Thus, by attaching the amino-terminal ubiquitin moiety to a protein in a way that makes its removal either impossible or inefficient, one can destabilize proteins by a distinct technique that is not directly based on the N-end rule.

In addition, variant cells can be developed which contain putative mutations in the "N-end" degrading protease(s) which either conditionally or nonconditionally stop degrading short-lived proteins. These cells can be used to overproduce proteins that ordinarily would be short-lived within the cell.

Figure 1 shows the construction of lacZ gene fusions.

Figure 2 shows experiments in which the half-lives of engineered β -gal proteins are directly measured.

Figure 3 shows the changing of amino acid residues at the ubiquitin- β gal junction (A) using the newly discovered properties of ubiquitin-specific processing protease and the amino acid sequence in the vicinity of the junction (B).

Figure 4 shows the presence of multiple ubiquitin moieties in metabolically unstable β -gal proteins.

Figure 5 shows a series of β -gal species containing ubiquitin in metabolically unstable β -gal proteins.

Figure 6 shows that both prokaryotic and eukaryotic long-lived intracellular proteins have stabilizing amino acid residues at their amino termini whereas secreted proteins exhibit a complementary bias.

Detailed Description of the Invention

The elucidation of the N-end rule is described in detail below. Briefly, this rule governing protein degradation was revealed by examining the in vivo half-lives of the enzyme β -galactosidase having various amino acid residues at its amino-terminus

-7-

and produced as a fusion protein with ubiquitin. When a chimeric gene encoding a ubiquitin- β -galactosidase fusion protein is expressed in the yeast *S. cerevisiae*, ubiquitin is cleaved off the nascent fusion protein, yielding a deubiquitinated- β -galactosidase (β gal). With one exception, this cleavage takes place regardless of the nature of the amino acid residue of β gal at the ubiquitin- β gal junction, thereby making it possible to expose selectively different residues at the amino-termini of otherwise identical β gal proteins. The β gal proteins so designed exhibited strikingly different half-lives in vivo, ranging from more than 20 hours to less than 3 minutes, depending upon the nature of the amino acid at the amino-terminus of β gal. Amino acids can be thus ordered according to the the half lives they confer on β gal when present at its amino-terminus. For example, the amino acids methionine, serine, alanine, threonine, valine, glycine and cysteine confer a half-life of more than 20 hours. Phenylalanine, leucine, asparagine, and lysine yield half lives of about three minutes. Arginine, the most destabilizing amino acid, confers a half life of about two minutes. (See Table 1 below for complete list of amino acids and the corresponding half-lives)

Currently known amino-terminal residues in long-lived, noncompartmentalized intracellular proteins from both prokaryotes and eukaryotes belong

-8-

virtually exclusively to the stabilizing class of amino acids, exactly as predicted by the N-end rule. This result strongly implicates the N-end rule in the selective degradation of intracellular proteins in general.

The appropriate amino-terminal amino acid appears to be an essential (though not necessarily a sufficient) requirement for the metabolic stability of a noncompartmentalized, intracellular protein. Thus, in order for a protein to be relatively stable intracellularly, a stabilizing amino acid should be present at the amino-terminus. The presence of a destabilizing residue at the amino-terminus of a protein is often, though not always, sufficient for its metabolic destabilization in vivo. When such destabilization occurs to a relatively small extent, further analysis shows either an insufficient accessibility of the amino-terminus or a lack of "permissible" sequence environment in the vicinity of the amino-terminus such as a lack of segmental mobility in the protein's amino-terminal region. The presence of a stabilizing amino acid at the amino-terminus at least in some cases (for instance, as observed for β -gal) will confer stability upon the protein. However, a stabilizing amino acid at the amino-terminus may not always confer a long half-life because other degradative pathways may be involved in determining the ultimate fate of protein. For example, endoproteolytic cleavages

-9-

(cleavages outside of terminal regions of the protein) may result in exposure of a destabilizing amino acid at the amino-terminus of a resulting product of the cleavage which is then rapidly degraded via the N-end rule pathway. Appropriate circumstances for use of a stabilizing amino acid can be ascertained empirically.

Although the N-end rule may be only one component (albeit a central one) of a more complex "half-life rule" which embraces other aspects of selective protein degradation in vivo, the N-end rule provides a rational, practicable approach for designing or changing protein structure in order to produce proteins which are more or less resistant to degradation by the N-end rule pathway than natural, unmodified protein. Proteins can be designed or modified at the protein or gene level to provide a desired amino acid of either the stabilizing or destabilizing class at their amino-terminus. The ability to regulate the half-life of a protein will allow one to modulate the intracellular activity of the protein.

A straightforward approach to modifying a protein to increase or decrease its metabolic stability is to directly engineer the amino-terminus of the protein at the protein level. To provide a desired amino-terminal amino acid, the amino-terminus of the protein of interest can be chemically altered, for example, by adding an amino acid of

the stabilizing or destabilizing class to the amino-terminus of a protein or polypeptide, employing an appropriate chemistry. Thus, for example, an unstable protein can be rendered more stable by adding a stabilizing amino acid residue (e.g. methionine, serine, alanine, threonine, valine, glycine or cysteine) to the amino-terminus of the protein. Conversely, a stable protein can be destabilized by adding a destabilizing amino acid to the amino-terminus. One distinct way to modify the amino-terminus of a protein would be to employ specific enzymes, amino acid-protein ligases, which catalyze posttranslational addition of a single amino acid to the protein amino-terminus. Other methods for nongenetic alterations of the same type can readily be ascertained by those skilled in the art.

In some proteins, the amino-terminal end is obscured as a result of the protein's conformation (i.e., its tertiary or quaternary structure). In these cases, more extensive alteration of the amino-terminus may be necessary to make the protein subject to the N-end rule pathway. For example, where simple addition or replacement of the single amino-terminal residue is insufficient because of an inaccessible amino-terminus, several amino acids (including lysine, the site of ubiquitin joining to substrate proteins) may be added to the initial

-11-

amino-terminus to increase the accessibility and/or segmental mobility of the engineered amino-terminus.

Modification or design of the amino-terminus of a protein can also be accomplished at the genetic level. Conventional techniques of site directed mutagenesis for addition or substitution of appropriate codons to the 5' end of an isolated or synthesized gene can be employed to provide a desired amino-terminal structure for the encoded protein. For example, so that the protein expressed has the desired amino acid at its amino-terminus the appropriate codon for a stabilizing amino acid can be inserted or built into the amino-terminus of the protein-encoding sequence.

At the same time, expressed proteins are often naturally modified within a cell after translation. This can include modification of the amino-terminus. For example, the amino-terminus can be acted on by an aminopeptidase which cleaves one or several amino acids from the amino-terminus. Amino acids may also be added to the amino-terminus by post-translational processing. This invention provides a way to "by-pass" still undefined rules of amino-terminal protein processing to expose exactly and specifically the desired amino acid residues at the amino-terminus of a mature processed protein species. To minimize the impact of such posttranslational events on the ultimate structure of the amino-terminus of a protein of interest, specific fusion proteins can be

-12-

designed wherein the amino-terminus of a protein of interest (designed to have the desired stabilizing or destabilizing structure) is preceded by a "masking" protein sequence fused to the amino-terminus. The fusion proteins are designed so that the protein sequence fused to the amino-terminus of the masking protein of interest is susceptible to specific cleavage at the junction between the two. Removal of the protein sequence thus unmask the amino-terminus of the protein of interest and the half-life of the released protein is thus governed by the predesigned amino-terminus. The fusion protein can be designed for specific cleavage in vivo, for example, by a host cell endoprotease or for specific cleavage in a in vitro system where it can be cleaved after isolation from a producer cell (which lacks the capability to cleave the fusion protein).

Ubiquitin is a broadly useful fusion partner for construction of a fused protein with a protein of interest: the discovery that artificial ubiquitin-protein fusions can be cleaved precisely by a cytoplasmic eukaryotic protease with little or no dependence on the protein to which ubiquitin is fused can be applied both in vivo and in vitro in protein engineering strategies, and is a major aspect of this invention. For example, the ubiquitin-protein fusion method can be used to artificially

-13-

generate authentic amino-termini in proteins produced through artificial means. Thus, amino-termini characteristic of natural eukaryotic or prokaryotic proteins can be generated by in vitro cleavage of ubiquitin-protein fusions produced in a prokaryotic host.

A specific methodology for producing ubiquitin- β -galactosidase fusion proteins is described in detail below. Genes encoding any other proteins can be substituted for LacZ (the β -gal gene) in this methodology.

In general, ubiquitin fusion proteins are expressed by a chimeric gene construct comprising, in 5' to 3' orientation, a ubiquitin gene linked to a gene encoding the protein of interest. The codon for the amino-terminal amino acid of the protein of interest is located immediately adjacent the 3' end of the ubiquitin gene. The fused gene product is cleaved endoproteolytically either in vivo or in vitro (using either pure or partially purified ubiquitin-specific protease identified in the present invention) at the junction between ubiquitin and the protein of interest to generate the protein of interest having the desired amino acid at its amino-terminus. There are a number of specific uses for the described ability to specifically engineer the protein's amino-terminus. One such use is established by the fact that the intracellular half-life of the released protein is governed by the principles of the N-end rule. Other applications of

the specific method for engineering the protein amino-terminus described herein range from adjusting the desired functional properties of a protein of interest, to modulating its antigenicity, and again, to other uses that can readily be ascertained by those skilled in the art.

This method of generating the desired amino acid residue at the amino-terminus of a protein of interest involves two novel components: one, the use of ubiquitin-protein fusions, and the other, the use of ubiquitin-specific processing protease that has been identified, and whose striking substrate requirements were discovered, in this work. Although the initial identification of the ubiquitin-specific protease has been made in vivo, the enzyme is also relatively stable and active in vitro (in extracts), and can readily be purified to homogeneity by techniques known to those skilled in the art. Furthermore, the substrate specificity of the ubiquitin-specific processing protease is highly conserved in evolution, being the same in yeast and mammals. The enzyme can be purified chromatographically from a crude extract by sequential chromatography on phosphocellulose, DEAE cellulose, and SH Sepharose among other methods known to those skilled in the art. Alternatively, the gene for this protease can be cloned by those skilled in the art. The cloned protease gene can be used either in vivo,

-15-

or, alternatively, the gene can be overexpressed in a suitable host, the overexpressed ubiquitin-specific protease purified and used for the same or similar purposes in vitro. The discovery of this enzyme, and detailed characterization of its substrate specificity herein provides for the in vitro and in vivo use of this enzyme.

The use of ubiquitin-protein fusions to allow the generation of a desired amino acid residue at the amino-terminus of a protein of interest can be extended to facilitate the purification of such proteins from producer cells. A gene can be readily constructed that encodes a convenient marker protein, such as streptavidin, linked to a ubiquitin-protein fusion construct described above. The resulting (marker protein)-ubiquitin-protein fusion can be simply isolated from producer cells by using the preselected property of the marker protein, for instance, the known ability of streptavidin to be isolatable by affinity chromatography on a biotin column. Thus, purified (marker protein)-ubiquitin-protein fusion can then be specifically cleaved by the ubiquitin-specific protease described in this invention to generate the final product, a protein of interest with the desired amino acid residue at its amino-terminus.

The codon for the amino-terminal amino acid of

mutagenesis techniques currently standard in the field. If the gene encoding the protein of interest is a synthetic gene the appropriate 5' codon can be built-in during the synthetic process. Alternatively, nucleotides for a specific codon can be added to the 5' end of an isolated or synthesized gene by ligation of an appropriate DNA sequence to the 5' (amino-terminus encoding) end of the gene.

Ubiquitin-like fusion partners capable of being cleaved by the ubiquitin-specific protease can also be used. In addition, fusion partners other than ubiquitin for masking the amino-terminus of a protein of interest can be used. In appropriate cases, the fusion proteins can be designed to contain a proteolytic cleavage site for a restriction endoprotease which has sufficiently narrow specificity so that only one target site is cleaved in a fusion protein. A crucial property of such a protease must be a sufficiently relaxed requirement for the nature of the amino acid residue(s) abutting the carboxy-terminal side of the cleavage site. The target site for cleavage is the junction between the fusion partner and the amino-terminus of the protein of interest and thus the recognition site for the endoprotease is located to provide for cleavage at this location. The commercially available protease, complement factor X_a , exhibits these properties and thus can be used to directly generate proteins with predetermined amino acid residues in the ultimate

-17-

position of their amino-termini (see, K. Nogai and H.C. Thogersen Nature 309:810 (1984)). The recognition site for the endoprotease can be engineered into the junction between the masking protein sequence and the 3' region encoding the amino-terminus of the protein of interest.

A different and distinct method for engineering short-lived proteins is provided in this invention by the discovery that ubiquitin-protein fusions, such as ubiquitin-Pro- β -galactosidase fusion (Table 1), that cannot be efficiently deubiquitinated are metabolically unstable. Thus, by attaching the amino-terminal ubiquitin moiety to a protein in a way that makes its removal either impossible or inefficient, one can destabilize a protein by a distinct technique which is qualitatively different from the method of generating the desired amino-terminus of a protein according to the requirements of the N-end rule. Prevention of the efficient deubiquitination of a ubiquitin-protein fusion can be achieved in several ways, for instance, by using a proline residue at the ubiquitin-protein junction as shown in Table 1, or by changing the amino acid sequence of ubiquitin near its carboxyl-terminus in such a way that the ubiquitin moiety is no longer recognized by the ubiquitin-specific processing protease but can still be recognized by the rest of the degradative pathway. These and other ways to reduce the rate of

deubiquitination of a ubiquitin-protein fusion can be readily ascertained by those skilled in the art.

The methods of this invention can be employed, inter alia, for regulating the half-life of a protein intracellularly. There are many instances where this capability is useful. For example, when a gene is introduced into a cell for expression therein, the expressed product can be designed for a long or short half-life depending upon the particular need.

In general, destabilized proteins which have short half-lives are more amenable to regulation of intracellular levels of the protein. The ability to finely regulate the intracellular levels and activity of a protein can be useful in therapy or in the work with in vitro cell cultures. In gene therapy, for example, a gene may be introduced into a cell to compensate for a genetic deficiency or abnormality. The gene can be inserted under control of an inducible promoter. Induction results in enhanced expression of the gene product and consequently, higher levels of the product within the cell. If the gene is designed to encode an unstable protein, the intracellular concentration of the expressed protein will be more quickly responsive to a later reduction in the rate of its synthesis because it does not persist within the cell. In this way, the intracellular level and/or activity of

-19-

the protein encoded by the inserted gene can be more finely regulated.

The method of this invention can also be used to expand the uses of selectable markers by shortening the time necessary for a phenotype related to the marker to become manifest. Toward this end, a product encoded by a marker gene can be destabilized by altering its amino-terminus according to the N-end rule. In this way selection for the negative phenotype can be facilitated because the product of the marker gene will be more quickly extinguished after function of the gene encoding the marker is abolished. An example is the thymidine kinase (tk) gene. The tk gene can be engineered to encode a less stable enzyme by introducing an appropriate destabilizing amino acid at the amino-terminus. Gene mutation resulting in tk⁻ phenotype will be more quickly manifested by cells because residual tk will be more quickly degraded. This can be especially useful in slow growing cells where more time is required to "dilute out" tk synthesized prior to transformation to the tk⁻ type.

The principles of protein modification based upon the N-end rule may also be employed in the design of cytotoxins. Proteinaceous cytotoxins can be designed as unstable proteins degradable by the N-end rule pathway so that they do not persist after their toxic action has been exerted on a target

-20-

cell. Reducing the lifetime of the toxin reduces the likelihood of killing nontargeted cells.

Discovery of the N-end rule pathway of degradation allows development of mutant cells having mutations in genes encoding essential components of the N-end rule pathway. For example, cells can be produced that either permanently or conditionally are unable to efficiently degrade otherwise short-lived proteins. These cells can be used to produce desired proteins that ordinarily would be unstable within a cell.

The invention is illustrated further by the following detailed description of the elucidation of the N-end rule.

Methods

Protein Sequencing

S. cerevisiae cells carrying pUB23 (Fig. 1), which encodes ub-Met- β gal (Fig. 3A), were labeled with [35 S] methionine, followed by extract preparation, immunoprecipitation of β gal and electrophoresis as described below. The wet polyacrylamide gel was subjected to autoradiography, the band of β gal was excised, and the electroeluted β gal was subjected to six cycles of radiochemical sequencing by Edman degradation. The sequencing was carried out by W. Lane at the MicroChem Facility of Harvard University.

-21-

Site-directed Mutagenesis

pUB23 (figure 1) was treated sequentially with Acc I, the Klenow fragment of pol I, and Bam HI. A fragment containing the Xho I site was purified and inserted between a filled-in Hind III site and a BAM HI site of the M13mp9 phage DNA. (J. Messing and J. Vieira, Gene 19, 263 (1982)). Site-directed mutagenesis (M. Smith, Annu. Rev. Genet. 19, 423 (1985)) was carried out as described by Kramer, W. et al. Nucl. Acids Res. 12, 9441 (1984) using a synthetic 25-residue oligodeoxyribonucleotide, containing ten bases on the 5' side and twelve bases on the 3' side of the Met codon of β gal. All four bases were allowed to occur at the original Met codon positions during synthesis. Primary phage plaques were screened by hybridization (Wood, N.I. et al. PNAS 82, 1585 (1985)), with the use of a 12-residue oligonucleotide probe spanning the region of codon changes and hybridizing to the original sequence. Nonhybridizing plaques containing inserts of the expected size were sequenced by the chain termination method. (Sanger, F. et al., PNAS 71 5463 (1977)). To transfer the desired constructs into the pUB23 background, replicative form DNA of mutant phages was digested with Xho I and Bam HI, and added to the same digest of the plasmid pLGSD5-ATG (see Fig. 1 and L. Guarente, Methods Enzymol., 101 181 (1983)). The ligated mixture was used to transform the E. coli strain MC1061. (M.J. Casadaban and S.N.

Cohen, J. Mol. Biol., 138 179 (1980)). Colonies containing plasmids of interest (in which the open reading frame of β gal had been restored) were recognized by their light blue color on X-gal plates.

Pulse-Chase Experiments

S. cerevisiae cells of the strain BWG-9a-1 (MAT his4 ura3 ade6), transformed (F. Sherman et al. Methods in Yeast Genetics Cold Spring Harbor Laboratory, N.Y., 1981)) with plasmids of interest were grown at 30°C to A₆₀₀ of approximately 5 in a medium of 2 percent galactose, 0.67 percent Yeast Nitrogen Base without amino acids (DIFCO), adenine (10 µg/ml) and amino acids including methionine (Sherman, F. et al., supra). Typically, cells from a 5 ml culture were harvested by filtration through the well of a Millipore microtiter filtration plate, washed several times on the filter with the same medium lacking methionine and resuspended in 0.3 ml of 1 percent galactose, 50 mM potassium phosphate buffer (pH 7.4). [³⁵S]methionine (50 to 100 µCi) was then added for 5 minutes at 30°C; the cells were collected by filtration and resuspended on 0.4 ml of the growth medium containing cycloheximide at 0.5 mg/ml. Samples (0.1 ml) were withdrawn at indicated times, and added to 0.75 ml of cold buffer A (see below for buffer composition) containing leupeptin, pepstatin A, antipain, aprotinin and chymostatin

-23-

(Sigma), (each at 20 µg/ml) in addition to 0.4 ml of glass beads. Immediately thereafter, the cells were disrupted by vortexing for approximately 3 minutes at 4°C; the extracts were centrifuged at 12,000g for 3 minutes and the radioactivity of acid-insoluble ³⁵S in the supernatants was determined. Aliquots of the supernatants containing equal amounts of the total acid-insoluble ³⁵S were processed for immunoprecipitation with a monoclonal antibody to gal. Ascitic fluid containing a molar excess of the antibody (at least tenfold) was added to each aliquot, with subsequent incubation at 4°C for 2 hours; protein A-Sepharose (Pharmacia) was then added, the suspension was incubated with rocking at 4°C for 30 minutes and centrifuged at 12,000g for 1 minute. The protein A-Sepharose pellets were washed three times in buffer A (see below) containing 0.1 percent sodium dodecyl sulfate (SDS), resuspended in an SDS, dithiotreitol (DTT)-containing electrophoretic sample buffer (U.K. Laemmli, Nature 227 680 (1970)), heated at 100°C for 3 minutes, and centrifuged at 12,000g for 1 minute. Equal aliquots of the supernatants were subjected to electrophoresis in a 7 percent discontinuous polyacrylamide-SDS gel (15 by 15 by 0.15 cm) with subsequent fluorography. In some experiments, the above protocol was not used, but the extracts were prepared by boiling cells directly in the presence of SDS, with essentially the same results.

Analysis of α - β gal proteins produced in *E. coli*.
Plasmid pUB23 (Figs. 1 and 3) was introduced into DS410, a minicell-producing *E. coli* strain. (N. Stoker, et al., in Transcription and Translation: A practical Approach B. D. Harnes and S. J. Higgins, Eds., IRL press, Oxford, 1984, p.153). Minicells were prepared and labeled for 60 minutes at 36°C with [³⁵S]methionine (600 Ci/mmol, Amersham) as described by N. Stoker et al., supra.

Labeled minicells were centrifuged, resuspended in 2 percent SDS, 10 mM DTT, 10 mM Na-HEPES (ph 7.5) and heated at 100°C for 3 minutes. After centrifugation at 12,000g for 1 minute the supernatant was diluted 20-fold with buffer A (1 percent Triton X-100, 0.15 M NaCl, 5 mM Na-EDTA, 50 mM Na-HEPES, ph 7.5), followed by the addition of phenylmethylsulfonyl fluoride (PMSF) and N-ethylmaleimide to 0.5 mM and 10mM, respectively. After 4 hours at 4°C, the sample was dialyzed against buffer A containing 0.5 mM PMSF overnight at 4°C, and processed for immunoprecipitation (as described above).

Analysis of ub- β gal proteins produced in yeast

S. cerevisiae cells carrying plasmids of interest were grown in 800 ml of a uracil-deficient medium, then harvested and disrupted with glass beads in buffer A containing leupeptin, pepstatin A, anti-pain, aprotinin and chymostatin (each at 3 μ g/ml). The extract was centrifuged at 12,000g for 3

-25-

minutes. Saturated ammonium sulfate was added to the supernatant to a final concentration of 57 percent. After overnight incubation at 4°C, the precipitated protein was collected by centrifugation at 23,000g for 30 minutes. The pellet was redissolved in buffer A containing protease inhibitors. After clarification at 12,000g for 3 minutes, the sample was passed through an affinity column which had been prepared by crosslinking an IgG fraction from an ascitic fluid (containing a monoclonal antibody to gal to Affi-Gel 10 (Bio-Rad)). The IgG fraction used for crosslinking had been purified from the ascitic fluid by affinity chromatography on protein A-Sepharose. After washing with buffer A lacking Triton X-100, the antibody-bound proteins were eluted with 0.25 M glycine-HCl (pH 2.6). The eluate was immediately adjusted to pH 7.5 with 1 M Na-HEPES (pH 8.5), and thereafter made 0.1 percent in SDS. The sample was concentrated by ultrafiltration in Centricon 30 (Amicon), and subjected to electrophoresis in a 7 percent discontinuous polyacrylamide-SDS gel (U.K. Laemmli, Nature (London) 227, 680 (1970)). Electroblothing of proteins to nitrocellulose, and immunoblot analysis with a peptide-mediated antibody to ubiquitin were performed as described by P.S. Swerdlow, D. Finley and A. Varshavsky, Analyt. Biochem. 156, 147. (1986). The same results were obtained with a different

-26-

antibody to ubiquitin obtained from A. Haas (Univ. of Milwaukee Med. School).

Detailed Description of the Figures

Figure 1 shows construction of a ubiquitin-lacZ gene fusion. PUB2, a pBR322-based genomic DNA clone (E. Ozkaynak, et al. Nature 312, 663 (1984) contains six repeats of the yeast ubiquitin-coding sequence (open boxes) together with the flanking regions (jagged lines). PUB2 was modified as shown in the diagram by placing a Bam HI site six bases downstream from the first ubiquitin repeat. This allowed the construction of an in-frame fusion (confirmed by nucleotide sequencing) between a single ubiquitin repeat and the lacZ gene of the expression vector pLGSD5-ATG (called G2 in L. Guarente, Methods Enzymol. 101 181 (1983)). The term "2 μ m" denotes a region of the pLGSD-ATG that contains the replication origin and flanking sequences of the yeast plasmid called 2 μ m circle (See L. Guarente, supra). Figure 3B shows the amino acid sequence of the fusion protein in the vicinity of the ubiquitin- β gal junction.

Figure 2 shows that the in vivo half-life of gal is a function of its amino-terminal residue. (lane a) Minicells isolated from an E. coli strain carrying PUB23, the initial ub-lacZ fusion (Figs. 1 and 3B), were labeled with [35 S]methionine for 60 minutes at 36°C, with subsequent analysis of β gal as

-27-

described. The same result was obtained when the labeled minicell SDS extract was combined with an unlabeled yeast SDS extract before immunoprecipitation of gal. (lane b) *S. cerevisiae* cells carrying PUB23 (Fig. 1), which encodes ub-Met- β gal (Fig. 3B), were labeled with [35 S]methionine for 5 minutes at 30°C, with subsequent analysis of β gal. The same result was obtained with the lengths of the [35 S]methionine labeling periods from 1 to 30 minutes, and with yeast extracts produced either by mechanical disruption of cells in the presence of protease inhibitors or by boiling the cells directly in an SDS-containing buffer. (lane c) Same as lane a but with *E. coli* cells carrying the control plasmid pLGSD5 (called G1 in L. Guarente, *supra*) which encodes gal. (lanes d to g) *S. cerevisiae* cells carrying PUB23 (Fig. 1), which encodes ub-Met- β gal (Fig. 3A), were labeled with [35 S]methionine for 5 minutes at 30°C (lane d) followed by a chase in the presence of cycloheximide for 10, 30, and 60 minutes (lanes e to g), extraction, immunoprecipitation, and analysis of β gal. (lanes h to j) Same as lanes d to f, but with ub-Ile- β gal (see Fig. 3A). (lanes k to m) Same as lanes h to j, but with ub-Gln- β gal. (lanes n to q) Same as lanes d to g, but with ub-Leu- β gal. (lanes r to u) Same as lanes d to g, but with ub-Arg- β gal. Designations: ori; origin of the separating gel; ub, ubiquitin; β gal, an electrophoretic band of the β gal protein.

-28-

containing a specified amino-terminal residue; in this terminology, the Met- β gal portion of ub-Met- β gal is designated as β gal. Arrowheads denote a metabolically stable, about 90kD degradation product of β gal which is formed apparently as the result of an in vivo endoproteolytic cleavage of a proportion of short-lived β gal proteins such as Leu- β gal and Arg- β gal (lanes n to u).

Figure 3 shows the changing amino acid residues of gal at the ubiquitin-gal junction. (A) The initial plasmid, pUB23 (Fig.1), which encodes ub-Met- β gal, was mutagenized as described above to convert the original Met codon ATG at the ub- β gal junction into codons specifying 19 amino acids other than Met. (The original round of mutagenesis shown in Figure 3, produced 15 out of 19 possible substitutions. The remaining four substitutions were produced later (see Table 1)). The arrowhead indicates the site of the deubiquitinating in vivo cleavage in the nascent fusion protein that occurs with all of the fusion proteins except ub-Pro-gal (see text). All of the constructions shown encode His as the second gal residue. In addition, in some of the constructions (ub-Met-His-Gly- β gal, ub-Met-Gln-Gly- β gal, and ub-Met-Gln-His-Gly- β gal, the last one produced by an insertion mutation; see Table 2), either His or Gln were following Met at the ubiquitin- β gal junction, with indistinguishable consequences for the metabolic stabilities of the

-29-

corresponding β gal proteins. (B) The amino acid sequence (in single-letter abbreviations) of ub-Met- β gal, the initial fusion protein (Fig. 1), in the vicinity of the ub- β gal junction. Single-letter amino acid abbreviations: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

Figure 4 shows that ubiquitin- β gal is short-lived if not deubiquitinated. (lanes a to g) *S. cerevisiae* cells carrying plasmids encoding ub-X- β gal fusion proteins in which X is the residue indicated at the top of each lane, were labeled for 5 minutes at 30°C with [³⁵S]methionine, followed by extraction, immunoprecipitation and analysis of β gal. Fluorographic exposures for these lanes were several times longer than those for similar patterns in Figure 2 to reveal the multiple ubiquitination of short-lived β gal proteins. (lanes h, i) Fluorographic overexposure of lanes n, o in Fig. 2 to reveal the "ladder" of multiply ubiquitinated Leu- β gal proteins in a pulse-chase experiment (zero and 10 minutes chase, respectively). (lane j) Same as lanes a to g, but with ub-Pro- β gal. (lane k) Same as lane j, but with ub-Gln- β gal. (lane l) Same as lane j. (lanes m to p) *S. cerevisiae* cells carrying a plasmid encoding ub-Pro- β gal were labeled for 5 minutes at 30°C with [³⁵S]methionine (lane m) followed by a chase in the presence of cycloheximide

upper small arrow to the right of lane p denotes ub-Pro- β gal, a small proportion of which is still present after 1 hour chase. The lower small arrow indicates an apparently deubiquitinated Pro- β gal that slowly accumulates during chase and is metabolically stable. The dot to the left of lane m denotes an endogenous yeast protein that is precipitated in some experiments by the antibody used. Square brackets denote the multiply ubiquitinated β -gal species (see Fig. 5). Other designation are as in Figure 2.

Figure 5 shows the "ladder" gal species containing ubiquitin. (lane a) S. cerevisiae cells carrying a plasmid which encodes ub-Gln- β gal, were grown and disrupted, and the extracts processed for isolation of β gal proteins by affinity chromatography on a column with immobilized antibody to β gal. The β gal proteins thus obtained were electrophoresed in a polyacrylamide-SDS gel, transferred to nitro-cellulose, and probed with an antibody to ubiquitin. (lane b) Same as lane a, but with ub-Pro- β gal. (lane c) Same as b but a longer autoradiographic exposure. (lane d) S. cerevisiae cells carrying a plasmid which encodes ub-Leu- β gal were labeled with [³⁵S]methionine for 5 minutes, with subsequent extraction, immunoprecipitation and electrophoresis of β gal (the same sample as in Figure 4, lane f). Square brackets denote the

-31-

multiply ubiquitinated Gln- β gal species detected with antibody to ubiquitin. The arrow indicates the band of ub-Pro- β gal, the initial fusion protein seen in lanes b and c. The arrowheads indicate the position of the band of deubiquitinated β gal (detectable by either Coomassie staining or metabolic labeling, but not with antibody to ubiquitin) derived from the ub-Gln- β gal fusion protein.

Figure 6 shows both prokaryotic and eukaryotic long-lived intracellular proteins have stabilizing amino acid residues at their amino-termini, whereas secreted proteins exhibit a complementary bias.

(A) 208 long-lived, directly sequenced, intracellular (noncompartmentalized) proteins with unblocked amino-termini from both prokaryotes (77 proteins) and eukaryotes (131 proteins) were distributed into three groups according to the nature of their amino-terminal residues as defined by the N-end rule (Table 1). All of the long-lived intracellular proteins examined bear exclusively stabilizing residues at their amino-termini. In panels B to D, analogous diagrams are presented for 243 secreted eukaryotic proteins (B), for 37 light and heavy immunoglobulin chains (C), and for 94 secreted eukaryotic toxins (D). Entries in C and D are subsets of entries in B. For proteins in B to D, the amino-termini compiled correspond, whenever the assignment is possible, to the most processed form of a protein that is still located within a

-32-

secreting cell. The data in A to D were manually compiled from the entire set of complete protein sequences available before 1981. The same conclusions have been recently reached after a more detailed and extensive, computer-assisted tabulation of protein amino-termini using the current National Biomedical Research Foundation database. The amino-terminal residues of Asn, Cys, His, and Trp were excluded from the compilation because in vivo half-lives of the corresponding gal proteins are still unknown (see, however, the legend to Table 1). Inclusion of the residues (Table 1) into a recently compilation of the same type did not change the original conclusion. Although the amino-terminal Pro was also excluded from the compilation, Pro appears to be a stabilizing residue for β gal (Table 1), consistent with the frequent presence of Pro at the amino-termini of long-lived noncompartmentalized proteins.

Results and Discussion

Rapid in vivo deubiquitination of a nascent ubiquitin- β gal fusion protein

Branched ubiquitin conjugates in which the carboxyl-terminal glycine of ubiquitin moieties is joined via an isopeptide bond to the ϵ -amino groups of internal lysine residues in proteins apparently

-33-

comprise the bulk of ubiquitin conjugates in eukaryotic cells. Joining of ubiquitin to the amino-terminal α -amino groups of target proteins, to yield linear ubiquitin conjugates, may also be chemically feasible. See A. Hershko, et al., PNAS USA 81: 7021 (1984). Whether or not linear ubiquitin-protein fusions are actually synthesized in vivo through posttranslational enzymatic conjugation of ubiquitin to protein amino-termini, such proteins can also be produced by constructing appropriate chimeric genes and expressing them in vivo. Construction of one such gene, which encodes yeast ubiquitin linked to gal of Escherichia coli, is shown in Figure 1.

When this gene is expressed in E. coli, the resulting β gal-containing protein has an apparent molecular mass which is approximately 6 kD greater than that of the control β gal, a value consistent with the presence of ubiquitin in the protein encoded by the chimeric gene. In contrast, when the same gene is expressed in yeast, the corresponding β gal protein is electrophoretically indistinguishable from the control β gal. This result is independent of the length of the [35 S]methionine labeling period (between 1 and 30 minutes). Furthermore, determination of the amino-terminal residue in the putative Met- β gal (half-life, $t_{1/2}$ 20 hours) by Edman degradation of the in vivo-labeled, gel-purified β -gal (Figure 2, lane d) directly confirmed the presence of the expected Met residue

-34-

(Figure 3A and Table 1) at its amino-terminus. Independent evidence that ubiquitin cleavage off the fusion protein occurs immediately after the last Gly residue of ubiquitin is presented below. We conclude that in yeast, ubiquitin is efficiently cleaved off the nascent ubiquitin- fusion protein, yielding a deubiquitinated β gal. [The absence of the deubiquitination reaction in *E. coli* is consistent with other lines of evidence indicating that prokaryotes lack both ubiquitin and ubiquitin-specific enzymes].

The ubiquitin- β -gal junction encoded by the chimeric gene, Gly-Met (Figures 1 and 3B), is identical to the junctions between adjacent repeats in the polyubiquitin precursor protein, which is efficiently processed into mature ubiquitin. Thus it is likely that the same protease, as yet uncharacterized biochemically, is responsible both for the conversion of polyubiquitin into mature ubiquitin and for the deubiquitination of the nascent ubiquitin- β gal protein. If so, one potential way to inhibit the in vivo deubiquitination of the ubiquitin- β gal (and thereby to allow analysis of metabolic consequences of a stable ubiquitin attachment to β -gal) would be to convert the Met residue of β -gal at the ubiquitin- β gal junction (Figure 3B) into other amino acid residues (Figure 3A). The unexpected results of such an approach are described below.

-35-

The in vivo half-life of β gal is a function of its amino-terminal residue. The ATG codon which specifies the original Met residue of gal at the ubiquitin junction (Figure 3B) was converted by sited directed mutagenesis into codons specifying 19 other amino acids (See Figure 3A and Table 1). These constructions differ exclusively in the first codon of β gal at the ubiquitin- junction (Figure 3A). After each of the 16 plasmids thus designed was introduced into yeast, analysis of the corresponding gal proteins pulse-labeled in vivo led to the following results (Figures 2, 4, and Table 1):

1) With one exception (see below), the efficient deubiquitination of the nascent ubiquitin- β gal occurs irrespective of the nature of the amino acid residue of β gal at the ubiquitin- β gal junction. Thus, the apparently ubiquitin-specific protease that cleaves the original ubiquitin- β gal protein at the Gly-Met junction is generally insensitive to the nature of the first residue of β gal at the junction (Figure 3A and Table 1). This result, in effect, makes it possible to expose different amino acid residues at the amino-termini of the otherwise identical β gal proteins produced in vivo.

2) The in vivo half-lives of the β gal proteins thus designed vary from more than 20 hours to less than 3 minutes, depending on the nature of the amino acid residue exposed at the amino-terminus of β gal (Figures 2, 4, and Table 1). Specifically,

-36-

deubiquitinated β gal proteins with either Met, Ser, Ala, Thr, Val, Cys or Gly at the amino-terminus have relatively long in vivo half-lives of 20 hours or more (figure 2, lanes d to g, and Table 1), similar to the half-life of a control β gal whose gene had not been fused to that of ubiquitin. In striking contrast, the β gal proteins with either Arg, Lys, Phe, Leu, Asp or Trp at the amino-terminus have very short half-lives, between approximately 2 minutes for Arg- β gal and approximately 3 minutes for Lys- β gal, Phe- β gal, Leu- β gal, Asp- β gal, Asn- β gal and Trp- β gal (Figure 2, lanes n to u, and Table 1). The half-life of β gal proteins with amino-terminal residues of either Gln, His or Tyr is approximately 10 minutes (Figure 2, lanes k to m, and Table 1), while an amino-terminal Ile or Glu confers on β gal a half-life of approximately 30 minutes (Figure 2, lanes h to j, and Table 1). Both pulse-chase and continuous labeling techniques were used in these experiments and yielded similar results.

The set of individual amino acids can be ordered with respect to the half-lives that they confer on β gal when exposed at its amino-terminus. The resulting rule (Table 1) is referred to as the "N-end rule".

-37-

Table 1: The N-end rule

In vivo			
Residue X in	Radius of gyration	deubiquitination	
ub-X- β gal	of X(A)	ub-X- β gal	$t_{1/2}$ of X- β gal
Met	1.80	+	
Ser	1.08	+	
Ala	0.77	+	
Thr	1.24	+	20 hours
Val	1.29	+	
Gly	0	+	
Cys		+	
Ile	1.56	+	
Glu	1.77	+	30 minutes
Tyr	2.13	+	
Gln	1.75	+	10 minutes
His			
Phe	1.90	+	
Leu	1.54	+	3 minutes
Trp		+	
Asp	1.43	+	
Asn		+	
Lys	2.08	+	

-38-

Arg	2.38	+	2 minutes

Pro	1.25	-*	7 minutes

*The rate of in vivo deubiquitination of ub-Pro- β gal is extremely low. The $t_{1/2}$ shown is that of the initial ub-Pro- β gal fusion protein (see Fig. 4, lanes j to p).

-39-

Legend to Table 1

The N-end rule. In vivo half-lives of β gal proteins in the yeast *S. cerevisiae* were determined either by the pulse-chase technique (for short-lived gal's; see below) or by measuring the enzymatic activity of gal in crude extracts. For the measurements of β gal activity, cells growing in a galactose-containing medium were transferred to an otherwise identical medium lacking galactose and containing 10 percent glucose. After further growth for at least 5 hours at 30°C, the ratio of β gal activities per cell before and after shift to glucose was determined for each of the β gal proteins. [GAL promoter-driven expression of the fusion genes (Figs. 1 and 3) is repressed in glucose medium]. For shorter-lived β gal proteins ($t_{1/2} < 1$ hour), the pulse-chase technique was used as well (Figs. 2 and 4). Electrophoretic bands of β gal proteins labeled with [25 S]methionine in pulse-chase experiments were cut out from scintillant-impregnated dried gels similar to those of Figs. 2 and 4, and 35 S in the bands was determined. The in vivo decay of short-lived gal proteins deviated from first-order kinetics in that the rate of degradation was lower when measured at later (1 hour) time points of the chase, the lower rate reflecting either a time-dependent toxic effect of cycloheximide or intrinsic characteristics of the in vivo degradation process. [Arrest of translation is required for an efficient short-term chase in *S. cerevisiae* because of the amino acid pool equilibration problems related to the presence of vacuoles in this organism]. The half-life values listed below were determined for the first 10 minutes of chase. Several lines of evidence (see description of Figs. 4 and 6) suggest that Pro is a stabilizing residue. The listed radii of gyration of amino acids are from. M. Levitt, *J. Mol. Biol.* 104:59 (1976).

-40-

Amino-terminal location of an amino acid is essential for its effect on gal half-life.

Site-directed mutagenesis was employed to insert a codon specifying a "stabilizing" amino acid (in this experiment, the Met residue) before the first codon of β gal at the ubiquitin- β gal junction (Table 2). Insertion of a stabilizing residue (Met) before either another stabilizing residue (Thr) or a variety of destabilizing residues (Gln, Lys, and Arg) at the ubiquitin- β gal junction invariably results in a long-lived deubiquitinated β gal (Table 2). Furthermore, in contrast to ubiquitin-Pro- β gal which is not only short-lived but also resistant to deubiquitination (Figure 4, lanes j to p, and Table 1), ubiquitin-Met-Pro- β gal is efficiently deubiquitinated in vivo to yield a long-lived Met-Pro- β gal (Table 2). These results show that both the identity of amino acid residue and its amino-terminal location (presumably the presence of a free α -amino group) are essential for its effect on β gal halflife. In addition, these results (Table 2) further support the expectation that ubiquitin cleavage off the fusion protein occurs immediately after the last Gly residue of ubiquitin (Figure 3A).

-41-

Table 2: N-terminal location of an amino acid is essential for its effect on β gal half-life	
Fusion protein	$t_{1/2}$ of deubiquitinated fusion protein
ub - Thr - β gal	>20 hours
ub - Met - Thr - β gal	>20 hours
ub - Gln - β gal	=10 minutes
ub - Met - Gln - β gal	>20 hours
ub - Lys - β gal	=3 minutes
ub - Met - Lys - β gal	>20 hours
ub - Arg - β gal	=2 minutes
ub - Met - Arg - β gal	>20 hours
ub - Pro - β gal	=7 minutes*
ub - Met - Pro - β gal	>20 hours

Amino-terminal location of an amino acid is essential for its effect on β gal half-life. The insertion mutants were obtained essentially as described for the initial set of mutants except that a 32-residue oligonucleotide, 5'-CCCGGGATCCGTGC(G/C/T/)(G/T)CATACCACCTCTTAG was used, containing 14 bases on the 5' side and 15 bases on the 3' side of the ambiguous codon inserted behind the Met codon. Bases in parentheses denote ambiguities at the positions 16 and 17 in the sequence. Half-lives of the corresponding β gal proteins were determined as described in the legend to Table 1.

-42-

A long-lived cleavage product of gal is formed during decay of short-lived β gal proteins.

The electrophoretic patterns of short-lived (but not of long-lived) β gal proteins invariably contain a specific, about 90 kD cleavage product of β gal (Figure 2, lanes n to u) which, unlike the parental β gal species, accumulates during the postlabeling (chase) period (Figure 4, lanes m-p). The 90 kD β gal fragment constitutes a relatively small proportion of the initial amount of the pulse-labeled β gal. Nonetheless, its existence implies that an in vivo endoproteolytic cleavage can rescue a protein fragment from the metabolic fate of its short-lived parental protein. It remains to be seen whether the resulting possibility of multiple half-lives within a single protein species is exploited in the design of naturally short-lived proteins.

Ubiquitin- β gal is short-lived when not deubiquitinated.

Ubiquitin-Pro- β gal, the only ubiquitin- β gal fusion that is not deubiquitinated in vivo (Fig. 4, lanes j to p), has a half-life of approximately 7 minutes (Table 1) which is less than 1 percent of the half-life of metabolically stable β gal proteins (Table 1). One interpretation of this result is that a metabolically stable ubiquitin attachment to protein amino-termini is sufficient to signal

-43-

degradation of acceptor proteins. This interpretation is consistent with earlier biochemical and genetic evidence that ubiquitination of short-lived proteins in a mammalian cell is essential for their degradation. At the same time, all ubiquitin- β gal fusion proteins other than ubiquitin-Pro- β gal are rapidly deubiquitinated in vivo (Table 1). Thus, the posttranslational amino-terminal ubiquitination of proteins may not be involved in an initial recognition or commitment step that designates proteins for degradation in vivo. Whether posttranslational amino-terminal ubiquitination (if it actually occurs in vivo) is essential for later stages of the degradation pathway remains to be determined. Earlier in vitro experiments indicated that preferential chemical modification of amino-termini of proteolytic substrates inhibits their degradation in an in vitro ubiquitin-dependent proteolytic system. Based on these data, it was proposed that amino-terminal ubiquitination of proteins is essential for their degradation. An alternative interpretation of the same results is that chemical blocking of proteins' amino-termini prevents the recognition of their amino-terminal residues by the "N-end rule" pathway whose initial stages are not necessarily ubiquitin-dependent.

(Fig. 2) reveal that the major band of a deubiquitinated, short-lived β gal protein coexists with a "ladder" of larger molecular mass, β gal-containing bands irregularly spaced at 4 to 7 kD intervals (Fig. 4, lanes c to g). No such larger species appear when the fluorograms of long-lived β gal proteins are similarly overexposed (Fig. 4, lanes a and b). Immunological analysis with both antibodies to β gal and antibodies to ubiquitin demonstrates that the "ladder" β gal species contain ubiquitin (Fig. 5).

A model for the selective degradation pathway.

With the exception of natural or engineered ubiquitin fusion proteins (Fig. 1 and Table 1), nascent proteins apparently lack ubiquitin moieties. The in vivo amino-terminal processing of nascent noncompartmentalized proteins generates their mature amino-termini via the action of amino-terminal peptidases whose substrate specificities have been partially characterized. (See Tsunasawa, S. et al. J. Biol. Chem. 260 5382 (1985); Boissel, J.P. et al. PNAS USA 82, 8448 (1985)). We suggest that the amino-termini thus generated are recognized by an "N-end-reading" enzyme. One specific model is that a commitment to degrade a protein molecule is made

-45-

as a result of the recognition of its amino-terminal residue by a stochastically operating enzyme whose probability of "clamping" at the target's amino-terminus is determined by the N-end rule (Table 1). Once the commitment is made, it is followed by a highly processive ubiquitination of the target protein which in the case of β gal is conjugated to more than 15 ubiquitin moieties per molecule of gal (Fig. 4, lanes c to g, and Fig. 5). The multiply ubiquitinated target protein is then degraded by a "down stream" enzyme (1) for which the ubiquitin moieties of the target serve as either recognition signals or denaturation (unfolding) devices, or both.

The ubiquitin-containing "ladder" β gal species (Fig. 4, lanes c to i, and Fig. 5) consist of apparently branched ubiquitin moieties joined to the ϵ -amino groups of internal lysine residues in β gal. Surprisingly, the "ladder" β gal species derived from ubiquitin-Pro- β gal are electrophoretically indistinguishable from the analogous species of β gal whose amino-terminal ubiquitin is cleaved off the nascent fusion protein (Fig. 4, lanes j to l, and Fig. 5). If the electrophoretically indistinguishable ubiquitinated β gal species are indeed structurally homologous, these results would be compatible with two alternative models in which, immediately after the first ubiquitins are branch-conjugated to β gal, either a branch-ubiquitinated ubiquitin-Pro- β gal

-46-

undergoes amino-terminal deubiquitination or, alternatively, and analogous *βgal* species lacking the amino-terminal ubiquitin moiety reacquires it. Experimental resolution of this ambiguity may establish whether the posttranslational amino-terminal ubiquitination of proteins (if it occurs *in vivo*) plays a role in the selective protein turnover.

Although both prokaryotic and eukaryotic proteins appear to follow the N-end rule (see below), bacteria apparently lack the ubiquitin system. Thus it is possible that the hypothetical N-end-recognizing protein is more strongly conserved between prokaryotes and eukaryotes than is the rest of the selective degradation pathway. Interestingly, the properties of a mammalian protein E3 whose presence is required for ubiquitination of proteolytic substrates by ubiquitin-conjugating enzymes *in vitro* are consistent with it being a component of the N-end-recognizing protein.

The N-end rule and the known amino-termini of intracellular proteins.

The unblocked amino-terminal residues in metabolically stable, noncompartmentalized proteins from both prokaryotes and eukaryotes are exclusively (Fig. 6A) of the stabilizing class (Met, Ser, Ala, Gly, Thr, Val), that is, the class that confers long *in vivo* half-lives on *gal* (Table 1). The one

-47-

short-lived intracellular protein for which the mature amino-terminus is known is the cII protein of phage lambda, the central component of a trigger that determines whether λ grows lytically or lysogenizes an infected cell. (Y.S. Ho, D. Wulff, M. Rosenberg, in Regulation of Gene Expression, I. Booth and C. Higgins, Eds. (Cambridge Univ. Press, London, 1986), p. 79; F. Banuett, M.A. Hoyt, L. McFarlane, H. Echols, I. Herskowitz, J. Mol. Biol. **187**, 213 (1986); M.A. Hoyt, D.M. Knight, A. Das, H.I. Miller, H. Echols, Cell **31**, 565 (1982); K. Nasmyth, Nature (London) **320**, 670 (1983)). The half-life of cII in lambda-infected E. coli is less than 3 minutes. Strikingly, the mature amino-terminus of cII starts with Arg (Ho, Y.W. et al., J. Biol. Chem. **257**, 9128 (1982)), the most destabilizing residue in the N-end rule (Table 1).

While the destabilizing amino acids can be either hydrophobic, uncharged hydrophilic or charged, they share the property of having larger radii of gyration than any of the stabilizing amino acids except Met (Table 1).

Amino-terminal residues in compartmentalized proteins are largely of the destabilizing class.

Figure 6 illustrates a striking difference between the choice of amino-terminal residues in long-lived, noncompartmentalized intracellular proteins (A) and in compartmentalized proteins, such

-48-

as secreted proteins (B), many of which are also long-lived in their respective extracellular compartments. One implication of this finding is that a single intracellular degradation pathway operating according to the N-end rule could be responsible both for the diversity of in vivo half-lives of intracellular proteins and for the selective destruction of compartmentalized proteins that are aberrantly introduced into the intracellular space. Some miscompartmentalized proteins may be more harmful to the cell than others. It is therefore of interest that secreted eukaryotic toxins contain strongly destabilizing residues (Arg, Lys, Leu, Phe, Asp) at their amino-termini more often than the general population of secreted proteins (Fig. 6, panels B to D).

The above consideration also suggest that, if the topological outside of a cell, such as lumens of the endoplasmic reticulum and Golgi, and the extracellular space, were to have degradation pathways analogous to the N-end rule pathway, they could be based on "inverted" versions of the N-end rule in which the amino-terminal residues that are destabilizing inside the cell are now the stabilizing ones and vice versa. Thus, the methods of the present invention should also be useful for manipulating the metabolic stability and other properties of compartmentalized proteins, including secreted ones.

-49-

Possible role of the N-end rule pathway in the turnover of long-lived proteins.

Long-lived intracellular proteins with destabilizing (Table 1) penultimate residues generally retain their initial amino-terminal methionine residue. The amino-terminal residues in long-lived intracellular proteins that do undergo amino-terminal processing are invariably of the stabilizing class (Table 1). An interesting possibility that would involve the N-end rule pathway in the turnover of long-lived proteins is that the rate-limiting step in the in vivo degradation of long-lived proteins may be a slow aminopeptidase cleavage that exposes a destabilizing residue, followed by rapid degradation via the N-end rule pathway. Note that fine-tuning of the rate of degradation may in this case be a function of the rate of aminopeptidase cleavage exposing a destabilizing residue rather than a function of the residue's destabilizing capacity according to the N-end rule.

The N-end rule and selective degradation of short-lived and damaged proteins.

The recognition of polypeptide chain folding patterns or of local chemical features that target an otherwise long-lived but damaged protein for selective degradation in vivo is unlikely to be

mediated directly by the N-end rule pathway. Instead, we suggest that specific proteases (analogous in function to nucleases that recognize specific lesions in DNA) cleave a targeted protein so as to expose a destabilizing residue at the amino-terminus of one of the two products of a cut. One testable prediction of this model is that the initial cleavage products of the degradation pathway should bear destabilizing residues at their N-termini. The preferential exposure of destabilizing residues at the amino-termini of products of the initial protein cleavages may be due either to intrinsic specificities of the proteases involved or simply to the fact that a majority of the amino acids belong to the destabilizing class (Table 1). Furthermore, initial cleavages of a protein would be expected to destabilize aspects of its original conformation, thus increasing the probability of further internal cuts. Whether the initial cleavage products of a protein would be degraded exclusively via the N-end rule pathway or would have to be processed further by additional internal cleavages should depend on several factors, such as the exposure of destabilizing residues at the amino-termini of initial cleavage products, and the relative rates of introduction of internal cuts. In this model, the N-end rule pathway should be essential for degradation of most of the metabolically unstable proteins, from chemically damaged,

-51-

prematurely terminated, improperly folded and miscompartmentalized ones to those that cannot assemble into native multisubunit aggregates, and finally to otherwise normal proteins that are short-lived in vivo. Thus, the metabolic instability of a protein may be mediated not only by the exposure of a destabilizing residue at its amino-terminus, but also by local conformational and chemical features of its polypeptide chain that result in proteolytic cleavages exposing destabilizing residues at the amino-termini of cleavage products.

For any given protein, a variety of factors in addition to the N-end rule may combine to modulate its half-life in vivo. Among such factors may be the flexibility and accessibility of the protein's amino-terminus (Thornton, J.M. and Sibanda, B.L., J. Mol. Bio. 167 443 (1983)), the presence of chemically blocking amino-terminal groups such as the acetyl group, the distribution of ubiquitinatable lysine residues near the amino-terminus, and other variables, such as the structure of the carboxy-terminus. Since amino-terminal regions of multisubunit proteins are commonly involved in the interfaces between subunits (Thornton, J.M. and Sibanda, B.L., J. Mol. Bio. 167 443 (1983)), quaternary structure of proteins is yet another parameter that is expected to modulate the impact of the N-end rule pathway on protein half-lives in vivo. Finally as

-52-

suggested above, the N-end rule pathway may also be essential for the degradation of proteins whose initial recognition as targets for degradation is independent of the structures at their amino-termini.

Functional significance of posttranslational addition of amino acids to amino-termini of proteins.

It has been known for many years that in both bacteria and eukaryotes there exists an unusual class of enzymes, aminoacyl-transfer RNA-protein transferases, which catalyze posttranslational conjugation of specific amino acids to the mature amino-termini of acceptor proteins in vitro (R.L. Soffer, in Transfer RNA: Biological Aspects, D. Soll, J.N. Abelson, P.R. Schimmel, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 1980), p493; C. Deutch, Methods Enzymol. 106, 198 (1984); A. Kaji, H. Kaji, G.D. Novelli, J. Biol. Chem. 240, 1185 (1965)). The posttranslational addition of amino acids to proteins in vivo dramatically accelerates in a stressed or regenerating tissue, for example, after physical injury to axons of nerve cells (S. Shyne-Athwal, R.V. Riccio, G. Chakraborty, N.A. Ingolia, Science 231, 603 (1986); N.A. Ingolia et al., J. Neurosci 3, 2463 (1983)). The N-end rule provides an explanation for this phenomenon. We suggest that selective changes in metabolic stability of otherwise undamaged, longlived proteins

-53-

that may be required by a changed physiological state of the cell are brought about by post-translational addition of destabilizing amino acids to the amino-termini of target proteins in vivo. Strikingly, the known reactions of posttranslational addition of amino acids to proteins (R.L. Soffer, in Transfer RNA: Biological Aspects, D. Soll, J.N. Abelson, P.R. Schimmel, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 1980), p493; C. Deutch, Methods Enzymol. 106, 198 (1984); A. Kaji, H. Kaji, G.D. Novelli, J. Biol. Chem. 240, 1185 (1965); S. Shyne-Athwal, R.V. Riccio, G. Chakraborty, N.A. Ingolia, Science 231, 603 (1986); N.A. Ingolia et al., J. Neurosci 3, 2463 (1983)) involve largely those amino acids (Arg, Lys, Leu, Phe, and Tyr) that are destabilizing according to the N-end rule (Table 1). Physiological states in which addition of destabilizing amino acids to proteins could be expected to occur include entry to and exit from the cell cycle, responses to chemical or physical stress, and specific differentiation events, such as erythroid differentiation and spermatogenesis, in which a proportion of preexisting, otherwise long-lived intracellular proteins is selectively degraded.

The in vitro degradation of some proteolytic substrates in a ubiquitin-dependent system from mammalian reticulocytes has recently been shown to depend on the presence of certain aminoacyl-tRNAs

-54-

(Ferber, S. and Crechanover, A., J. Biol. Chem. 261 3128 (1986)). We suggest that this phenomenon also reflects a requirement for posttranslational addition of specific destabilizing amino acids to the amino-termini of proteolytic substrates. The initial proteolytic substrates in question have amino-terminal residues of Asp or Glu, both of which are destabilizing according to the N-end rule (Table 1). This raises an interesting and testable possibility that certain amino-terminal residues in proteins may not be directly destabilizing as such but only through their ability to be conjugated to other destabilizing residues.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims:

-55-

CLAIMS

1. A method of regulating the stability of an intracellular protein, comprising modifying the amino-terminus of a protein to provide for exposure at the amino-terminus of a desired amino acid of the stabilizing or destabilizing class according to the N-end rule of protein degradation.
2. A method of destabilizing an intracellular protein, comprising modifying the protein to provide an amino-terminal amino acid of the destabilizing class.
3. A method of Claim 2, wherein an amino acid of the destabilizing class is selected from the group consisting of Isoleucine, Glutamic acid, Tyrosine, Glutamine, Phenylalanine, Leucine, Asparagine, Lysine or Arginine.
4. A method of producing a protein which is more or less resistant to degradation than its wild-type counterpart, comprising introducing into the amino-terminus of the wild-type protein a desired amino acid of the stabilizing or destabilizing class according to the N-end rule of protein degradation to produce a

protein is produced as a fusion protein containing a masking protein joined to its amino-terminus, the masking protein being cleavable at the junction with the amino-terminus of the modified protein.

6. A method of Claim 5, wherein the masking protein is ubiquitin.
7. A method of Claim 6, wherein ubiquitin-specific protease is used to cleave ubiquitin from the fusion protein at the ubiquitin-protein junction either in vivo or in vitro.
8. A method of Claim 4, wherein the masking protein is joined to the amino-terminus of the modified protein through a protein sequence comprising a site for recognition by an endoprotease.
9. A method of Claim 8, wherein the endoprotease is complement factor X_a .
10. A method of producing a metabolically unstable protein comprising producing the protein as a

-57-

fusion protein wherein ubiquitin is joined to the amino-terminus of the protein in a manner such that the fusion protein cannot be efficiently deubiquitinated.

11. A method of Claim 10, wherein the fusion protein comprises ubiquitin-proline-protein.
12. A method of Claim 10, wherein the fusion protein contains ubiquitin having a modification in the amino acid sequence recognized by ubiquitin-specific processing protease so as to reduce the efficiency of deubiquitination of the fusion protein by the protease.
13. A gene construct encoding a fusion protein, the gene construct comprising a DNA sequence encoding a masking protein joined to a DNA sequence encoding a protein of interest having a predetermined amino acid of the stabilizing or destabilizing class at its amino-terminus, the masking protein being cleavable at the junction with the amino-terminus of the protein of interest such that cleavage results in exposure of the amino-terminus of the protein of interest.
14. A method of Claim 13, wherein the masking protein is ubiquitin.

-58-

15. A method of producing a protein having a predetermined amino-terminal structure, comprising expressing the protein as a fusion protein wherein the amino-terminus of the protein is fused to a masking protein which is specifically cleavable in vivo or in vitro at the junction with the amino-terminal amino acid of the protein.
16. A method of Claim 15, wherein the masking protein is cleavable by an endoprotease.
17. A method of Claim 16, wherein the masking protein is ubiquitin and the cleaving endoprotease is ubiquitin-specific protease.
18. A method of producing a protein having a predetermined amino-terminal structure, comprising:
 - a. preparing a DNA construct encoding a fusion protein comprising:
 - i) a structural gene encoding the protein, the gene having a sequence at its 5' end encoding the predetermined amino-terminal structure of the protein; and
 - ii) DNA encoding a masking protein linked to the 5' end of the structural gene encoding the protein, the masking

-59-

- protein being specifically cleavable in vivo or in vitro at its junction with the amino-terminus of the protein; and
- b. expressing the DNA construct in a host cell to produce the fusion protein encoded by the DNA construct.
19. A method of Claim 18, wherein the fusion protein is cleaved within the host cell to release the protein having the predetermined amino-terminal structure.
20. A method of Claim 18, wherein the fusion protein is produced in a cell which lacks the capability to cleave the masking protein and the fusion protein is cleaved in vitro to release the protein having the predetermined amino-terminal structure.
21. A method of Claim 18, wherein the masking protein is ubiquitin.
22. A DNA construct for producing a protein having a predetermined amino-terminal structure, comprising:
- a. a structural gene encoding the protein, the gene having a sequence at its 5' end

encoding the predetermined amino-terminal structure of the protein; and

- b. DNA encoding a masking protein linked to the 5' end of the structural gene encoding the protein, the masking protein being specifically cleavable in vivo or in vitro at its junction with the amino-terminus of the protein.
23. A method of Claim 22, wherein the masking protein is ubiquitin.
24. Isolated ubiquitin-specific protease capable of cleaving ubiquitin from the amino-terminus of a protein to which it is joined at the junction between ubiquitin and the amino-terminus of the protein.
25. An isolated DNA sequence encoding ubiquitin-specific protease of Claim 24.
26. A mutant cell line which is conditionally or unconditionally incapable of synthesizing one or more proteases involved in the N-end rule degradative pathway.

1/6

FIGURE 1

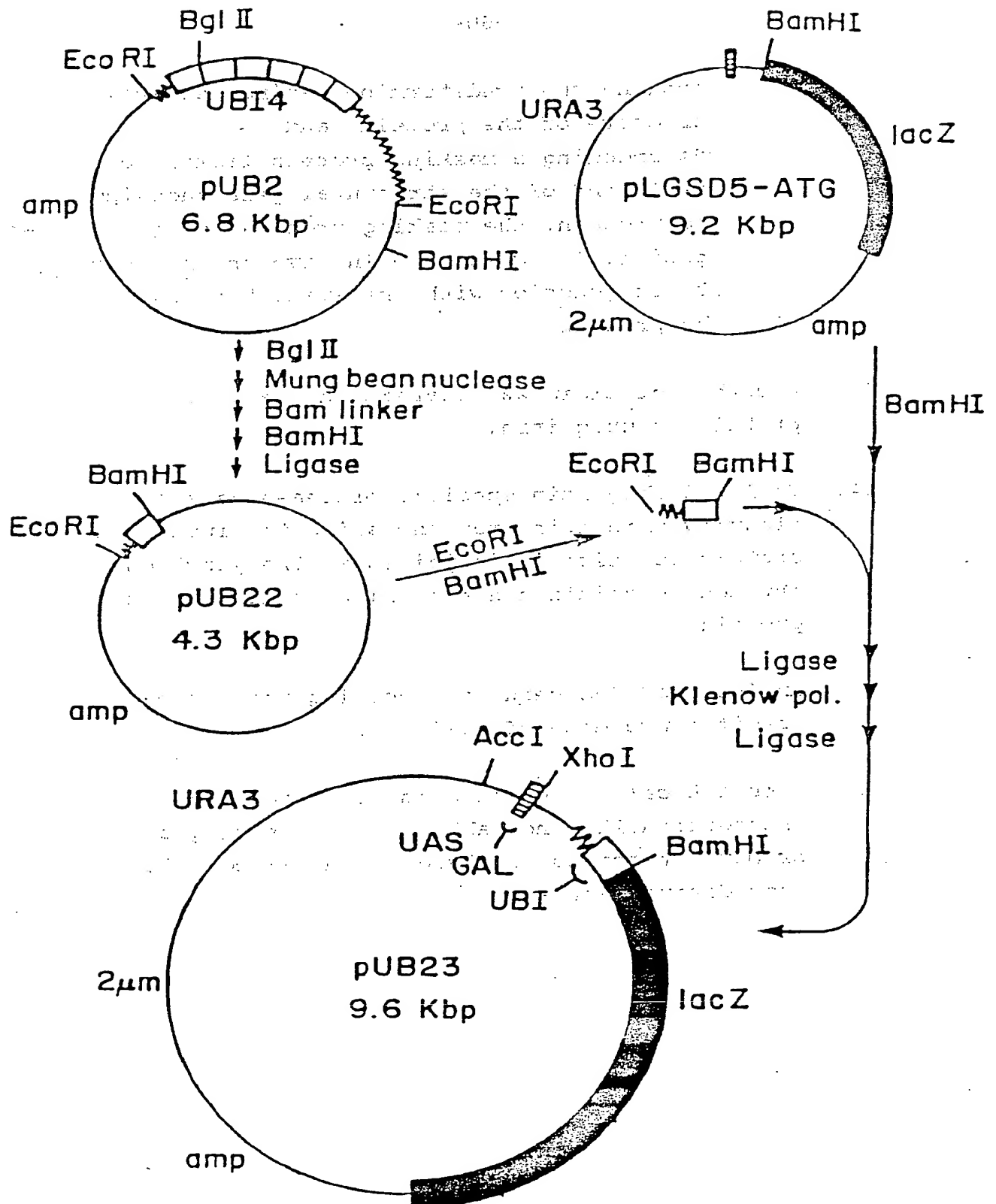
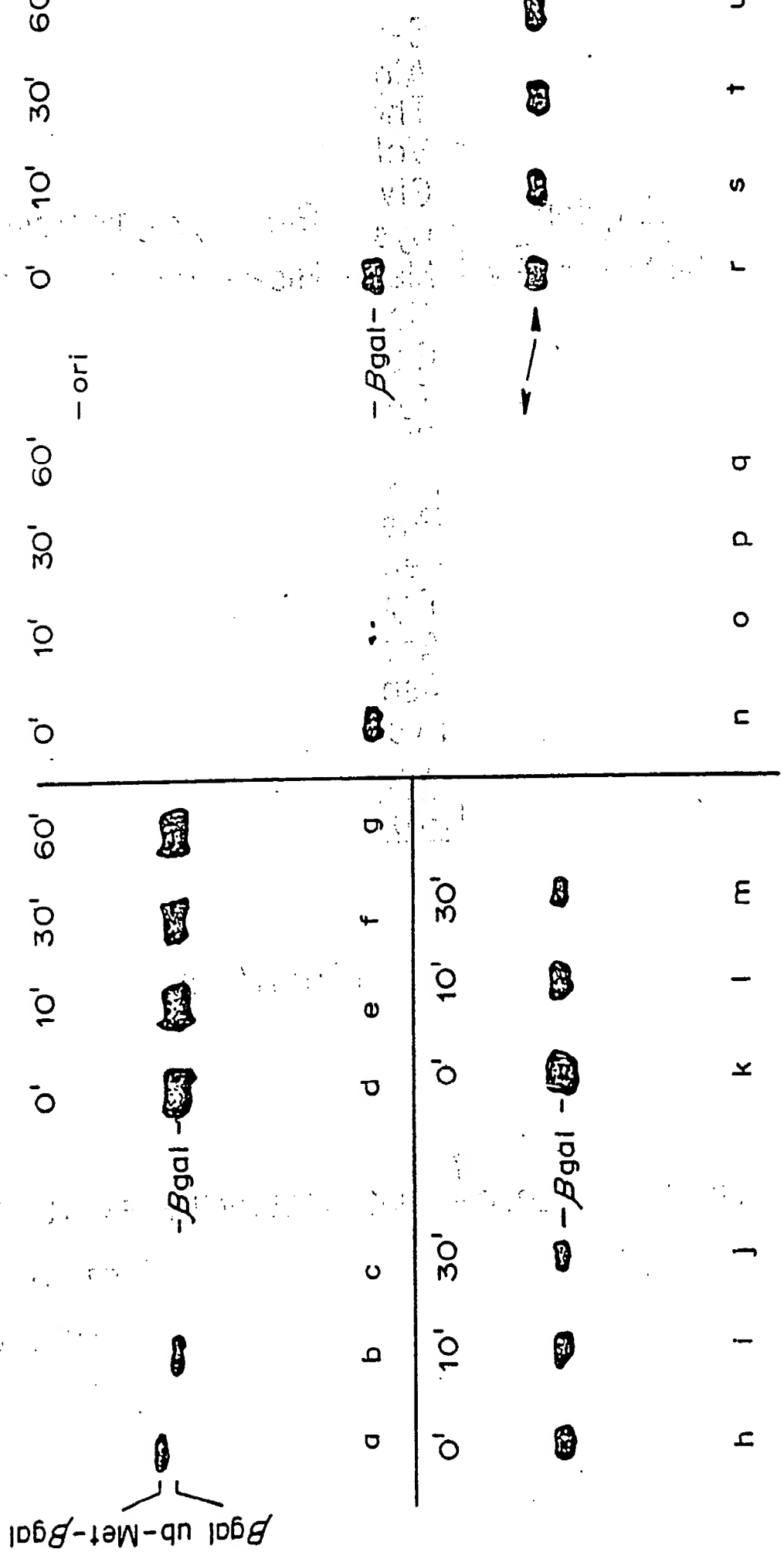


FIGURE 2



3/6

FIGURE 3a

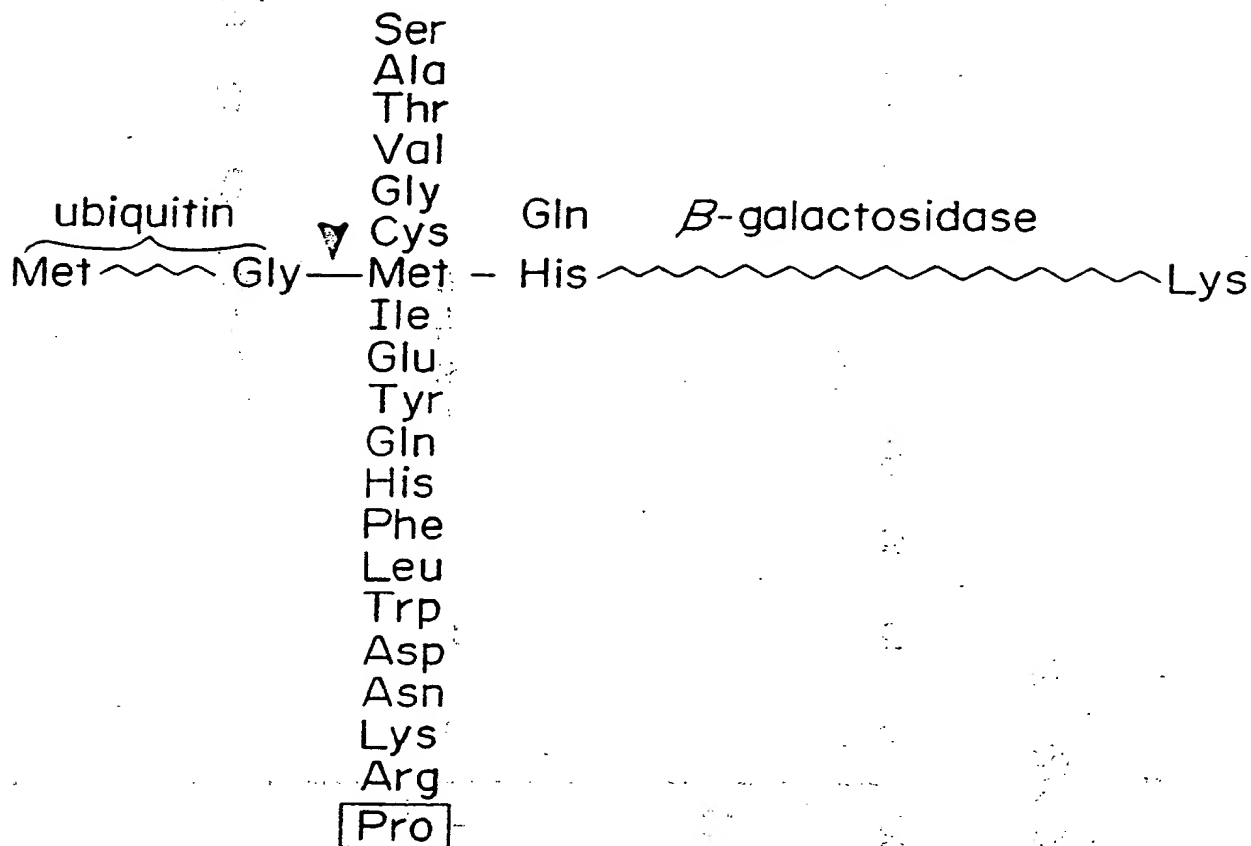


FIGURE 3b

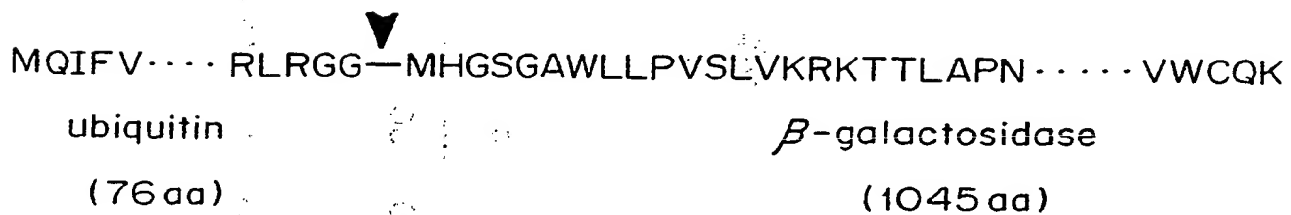
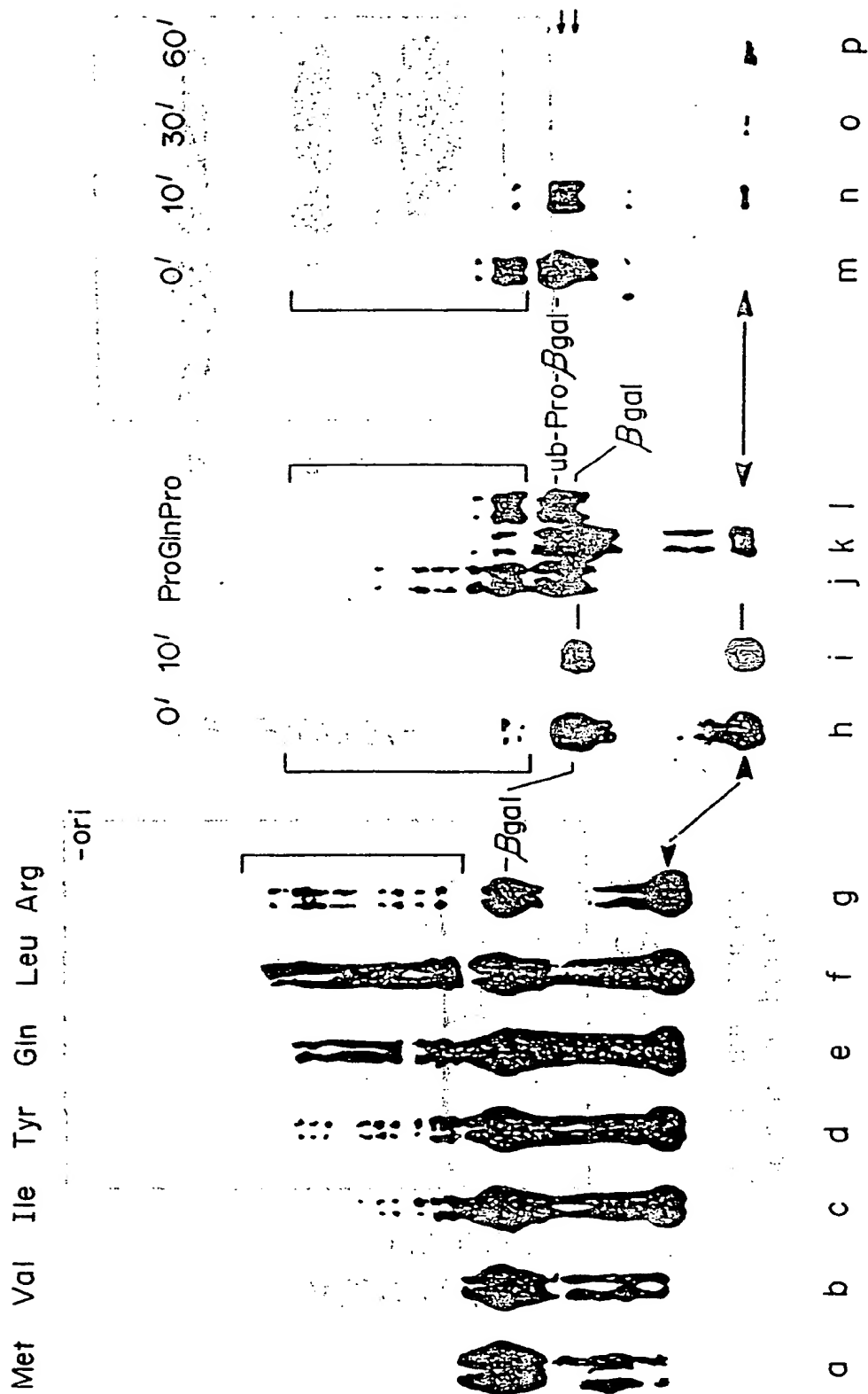


FIGURE 4



5/6

FIGURE 5

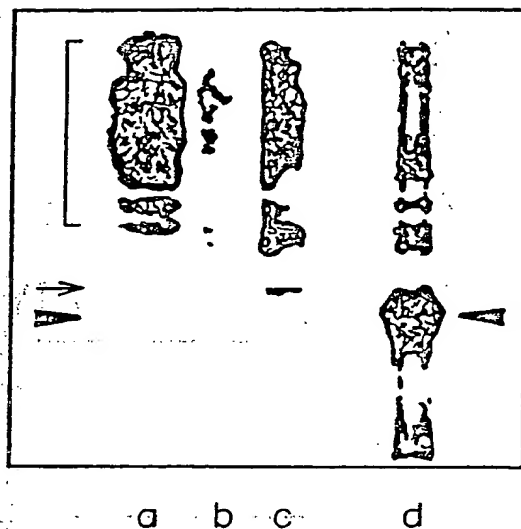
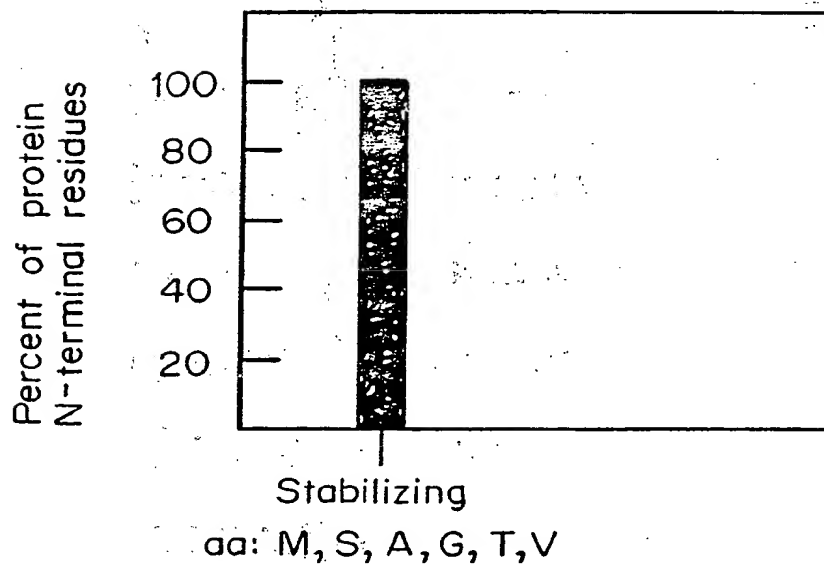


FIGURE 6a



SUBSTITUTE SHEET

6/6

FIGURE 6b

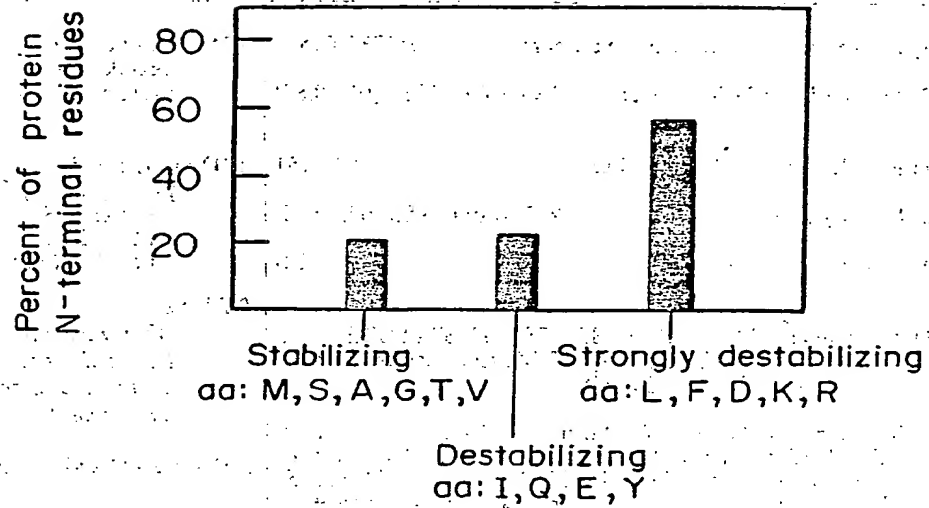


FIGURE 6c

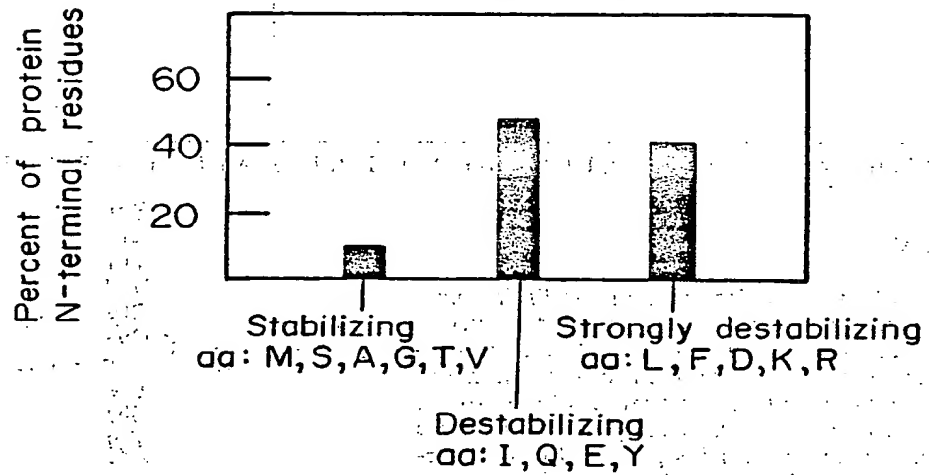
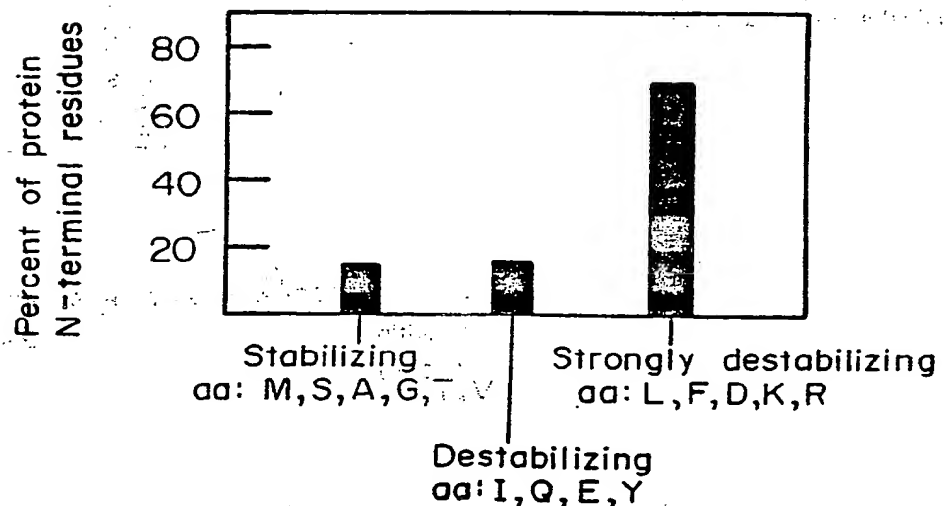


FIGURE 6d



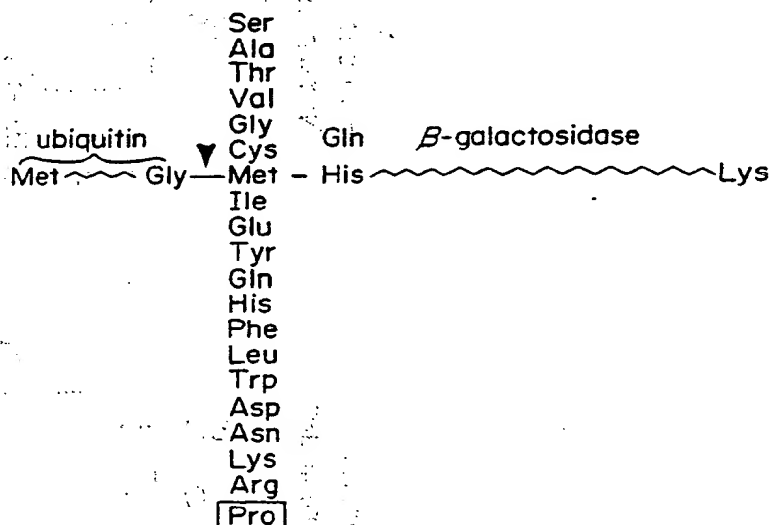


INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification⁴ : C12N 15/00, C12P 21/00, 19/34 C12N 9/60	A3	(11) International Publication Number: WO 88/ 02406 (43) International Publication Date: 7 April 1988 (07.04.88)
(21) International Application Number: PCT/US87/02522 (22) International Filing Date: 1 October 1987 (01.10.87) (31) Priority Application Number: 915,151 (32) Priority Date: 2 October 1986 (02.10.86) (33) Priority Country: US (71) Applicant: MASSACHUSETTS INSTITUTE OF TECHNOLOGY [US/US]; 77 Massachusetts Avenue, Cambridge, MA 02139 (US). (72) Inventors: BACHMAIR, Andreas ; 14 Bidelow Street, Cambridge, MA 02139 (US). FINLEY, Daniel ; 27 Lee Street, Cambridge, MA 02139 (US). VARSHAVSKY, Alexander ; 311 Commonwealth Avenue #30, Boston, MA 02115 (US).	(74) Agents: DeCONTI, Giulio, A., Jr. et al.; Hamilton, Brook, Smith & Reynolds, Two Militia Drive, Lexington, MA 02173 (US). (81) Designated States: AT (European patent), BE (European patent), CH (European patent), DE (European patent), FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i> (88) Date of publication of the international search report: 14 July 1988 (14.07.88)	

(54) Title: METHODS OF REGULATING METABOLIC STABILITY OF PROTEINS**(57) Abstract**

Methods of designing or modifying protein structure at the protein or genetic level to produce specified amino termini *in vivo* or *in vitro*. The methods can be used to alter the metabolic stability and other properties of the protein or, alternatively, to artificially generate authentic amino-termini in proteins produced through artificial means. The methods are based upon the introduction of the use of artificial ubiquitin-protein fusions, and the discovery that the *in vivo* half-life of a protein is a function of the amino-terminal amino acid of the protein.



MQIFV...RLRGG—MHGSGAWLLPVSLVKRKTTLAPN...VWCQK
 ubiquitin (76 aa) β-galactosidase (1045 aa)


FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT:

AT Austria	FR France	ML Mali
AU Australia	GA Gabon	MR Mauritania
BB Barbados	GB United Kingdom	MW Malawi
BE Belgium	HU Hungary	NL Netherlands
BG Bulgaria	IT Italy	NO Norway
BJ Benin	JP Japan	RO Romania
BR Brazil	KP Democratic People's Republic of Korea	SD Sudan
CF Central African Republic	KR Republic of Korea	SE Sweden
CG Congo	LI Liechtenstein	SN Senegal
CH Switzerland	LK Sri Lanka	SU Soviet Union
CM Cameroon	LU Luxembourg	TD Chad
DE Germany, Federal Republic of	MC Monaco	TG Togo
DK Denmark	MG Madagascar	US United States of America
FI Finland		

INTERNATIONAL SEARCH REPORT

International Application No PCT/US 87/02522

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) *		
According to International Patent Classification (IPC) or to both National Classification and IPC		
IPC ⁴ : C 12 N 15/00; C 12 P 21/00; C 12 P 19/34; C 12 N 9/60		
II. FIELDS SEARCHED		
Minimum Documentation Searched ⁷		
Classification System	Classification Symbols	
IPC ⁴	C 12 N	
Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched *		
III. DOCUMENTS CONSIDERED TO BE RELEVANT ⁹		
Category *	Citation of Document, ¹¹ with indication, where appropriate, of the relevant passages ¹²	Relevant to Claim No. ¹³
X	Nucleic Acids Research, volume 13, no. 4, February 1985, IRL Press Ltd, (Oxford, GB), B. Nilsson et al.: "Efficient secretion and purification of human insulin-like growth factor I with a gene fusion vector in Staphylococci", pages 1151-1162 see page 1153, line 20 - page 1158, line 20	1-5,13
Y	--	6-9,14
X	WO, A, 85/04186 (BIOGEN) 26 September 1985 see page 8, lines 5-15; page 11, line 32 - page 12, line 16; claims 1,2	1,2
Y	Nature, volume 309, 28 June 1984, K. Nagai et al.: "Generation of β -globin by sequence-specific proteolysis of a hybrid protein produced in Escherichia coli", pages 810-812 see the whole article cited in the application	7-9
<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>* Special categories of cited documents: ¹⁰</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> </div> <div style="width: 45%;"> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"&" document member of the same patent family</p> </div> </div>		
IV. CERTIFICATION		
Date of the Actual Completion of the International Search		Date of Mailing of this International Search Report
25th February 1988		09 JUN 1988
International Searching Authority		Signature of Authorized Officer
EUROPEAN PATENT OFFICE		 P.C.G. VAN DER PUTTEN

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
A	Science, volume 198, no. 4321, 9 December 1977, K. Itakura et al.: "Expression in Escherichia coli of a chemically synthesized gene for the hormone somatostatin", pages 1056-1063 see the whole article	
X	WO, A, 85/00831 (AMGEN) 28 February 1985 see claims 8-10,22,23	1-5,13
Y		6-9,14
Y	Proc. Natl. Acad. Sci, USA, volume 81, November 1984, A. Hershko et al.: "Role of the α -amino group of protein in ubiquitin-mediated protein break- down", pages 7021-7025 cited in the application	6-9,14
P,X	Science, volume 234, no. 1883, 10 October 1986, A. Bachmair et al.: "In vivo half-life of a protein is a function of its amino-terminal residue", pages 179-186 see the whole article	1-9,13,14

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

V. ☐ OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE ¹

This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1. ☐ Claim numbers..... because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claim numbers..... because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claim numbers..... because they are dependent claims and are not drafted in accordance with the second and third sentences of PCT Rule 6.4(b).

VI. ☒ OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING ²

This International Searching Authority found multiple inventions in this international application as follows:

1. Claims 1-9, 13, 14

5. Claim 26

2. Claims 10-12

3. Claims 15-23

4. Claims 24-25

Please refer to Form PCT/ISA/206
dated 8 March 1988.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.

2. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

3. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

1-9, 13, 14

4. ☐ As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not invite payment of any additional fee.

Remark on Protest

☐ The additional search fees were accompanied by applicant's protest.

☐ No protest accompanied the payment of additional search fees.

**ANNEX TO THE INTERNATIONAL SEARCH REPORT
ON INTERNATIONAL PATENT APPLICATION NO.**

US 8702522
SA 19180

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report. The members are as contained in the European Patent Office EDP file on 27/05/88. The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A- 8504186	26-09-85	EP-A- 0174999 JP-T- 61501430	26-03-86 17-07-86
WO-A- 8500831	28-02-85	EP-A- 0135094 JP-T- 60501989	27-03-85 21-11-85